Probabilistic Approaches to Energy Systems

Iversen, Jan Emil Banning

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Probabilistic Approaches to Energy Systems

February 9, 2015

Emil B. Iversen

Kongens Lyngby 2015
IMM-PhD-2015–363

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Building 303B, Matematiktorvet, DK-2800 Kongens Lyngby, Denmark
Phone +45 45 25 30 75
jebi@dtu.dk
www.compute.dtu.dk

PhD Thesis:
ISBN xxx-xx-xxx-xxxx-x,
IMM-PhD-2015–363,
ISSN 0909-3192

Preface

This thesis was prepared at the Department of Applied Mathematics and Computer Science at the Technical University of Denmark (DTU) in partial fulfillment of the requirements for acquiring a Ph.d. degree. The project was funded jointly by DTU and Det Strategiske Forskningsråd through the Ensymora project (no. 10-093904/DSF).

The thesis concerns the development of probabilistic methods for forecasting and optimization to support the integration of renewable energy sources in the energy system. In particular the thesis focuses on using continuous time methods and of exploiting the inherent benefits thereof.

The thesis consists of a summary report and a collection of 7 papers, documenting the work that was carried out from October 2011 trough January 2015. Four of these papers are either published or in press in international peer reviewed journals and three are currently submitted to international journals.

Nørrebro, January 2015

E. B. Iversen

Acknowledgements

This thesis presents the essence of a very intense period spanning more than three years where I enjoyed the support of many people. This thesis would have looked very different had I not received the personal and intellectual support that I enjoyed over these years. I would like to express my gratitude to some of these people here.

First and foremost, I would like to thank my three supervisors, prof. Juan Miguel Morales, prof. Jan Kloppenborg Møller and prof. Henrik Madsen. I have learned so much from their supervision. I would like to thank them for their unrelenting support.

I would also like to acknowledge Prof. Jan Kleissl for hosting me during my stay at University of California, San Diego and for co-authoring a paper. I learned a lot during my stay and had an incredible experience.

I would like to thank my colleagues at DTU compute for contributing to an inspiring work environment and their willingness to engage in academic discussions and assist with technical issues ranging from printing to coding.

Finally I want to thank my family and friends for making life much more enjoyable outside statistics and modeling. A special thanks goes to Lærke who, with her unrelenting support and encouragement, made both the good and the not so good times better.

Summary

Energy generation from wind and sun is increasing rapidly in many parts of the world. This presents new challenges on how to integrate this uncertain, intermittent and non-dispatchable energy source. This thesis deals with forecasting and decision making in energy systems with a large proportion of renewable energy generation. Particularly we focus on producing forecasting models that can predict renewable energy generation, single user demand, and provide advanced forecast products that are needed for an efficient integration of renewable energy into the power generation mix. Such forecasts can be useful on all levels of the energy systems, ranging from the highest level, where the transmission system operator is concerned with minimizing system failures and is aided by wind power forecasts, to the end user of energy where power price forecasts are useful for users with flexible power demand.

The main contributions of this thesis lie in the realm of using gray box models to produce forecasts for energy systems. Gray box models can be defined as a crossover between physical models (or white box models), that base their model on a physical understanding of the system at hand, and data driven models (or black box models) that focus on accurately describing the data without considering physical limitations of the system. Integrating these physical structures into a data driven approach allows for producing better forecasts with more accurate predictions. In this thesis we have developed and applied methodologies for gray box modeling to produce forecasts for vehicle driving patterns, solar irradiance, wind speeds, wind power, and solar power. The model for driving patterns has subsequently been used as input into an optimization algorithm for charging a single electric vehicle. In a subsequent study the behavior of a fleet of electric vehicles has been studied.

In the thesis we go through various examples of forecasts products and their applications. We emphasize that forecasting can not stand alone and should be complimented

by optimization and decision making tools for an efficient integration of renewable energy. Thus forecast products should be developed in unison with the decision making tool as they are two sides of the same overall challenge.

Dansk Resumé (Danish Summary)

Energiproduktion fra vind og sol er hastigt voksende i store dele af verden. Dette giver nye udfordringer indenfor integrering af denne usikre, variable og ikke styrbare energikilde. Denne afhandling omhandler prognoser og beslutningsproblemer for energisystemer med en stor andel af vedvarende energi. Især fokuserer vi på at producere prognosemodeller, der kan forudsige produktionen vedvarende energi og enkelt bruger efterspørgsel, samt producere de avancerede forudsigelsesprodukter, der er nødvendige for en effektiv integration af vedvarende energi i energiforsyningen. Sådanne prognoser kan være nyttigt på alle niveauer i energisystemerne. Der spænder fra det højeste niveau, hvor distributionssystemoperatøren søger at minimering systemnedbrud og bliver hjulpet af vindkraft prognoser til at forudsiger behovet for fleksible reserver, til slutbrugeren af energi, hvor prognoser for elprisen kan være nyttige for bedst muligt at time dennes elforbrug.

De vigtigste bidrag fra denne afhandling ligger inden for brugen af *gray box* modeller til at producere prognoser for energisystemer. *Gray box* modeller kan defineres som en blanding mellem fysisk modeller (eller *white box* modeller), som baserer sig på en fysisk forståelse af det pågældende systemet, og data drevene modeller (eller *black box* modeller), der fokuserer på præcist at beskrive data uden at tage det aktuelle fysiske systems begrænsninger i betragtning. Integration af disse fysiske strukturer i en data drevet model giver mulighed for at producere bedre prognoser med mere præcise forudsigelser. I denne afhandling har vi udviklet og anvendt metoder indenfor *gray box* modellering til at forudsige på kørselsmønstre, solens indstråling, vindhastigheder, vindkraft og solenergi. Kørselsmønstrene er efterfølgende anvendt som input til en optimerings algoritme for opladning af en enkelt elbil. Herefter er betydningen af

variable elpriser på omkostninger ved at oplade en elbil blevet studeret.

I afhandlingen gennemgår vi forskellige eksempler på forudsigelsesprodukter og deres anvendelse inden for energisystemer. Vi understreger, at prognoser ikke kan stå alene og bør komplementeres af optimering og beslutningsværktøjer for en effektiv udnyttelse af vedvarende energi. Således skal forudsigelsesprodukterne udvikles i konsensus med de aktuelle beslutningsværktøj, da de er to delløsninger af det samme overordnede problem.

List of Publications

International Journals

The following papers were published or submitted for publication in international journals during the project period. They constitute the main contributions of the Ph.d. project and we advice the reader to pick up on the specific details in the papers after reading the summary report.

[A] E. B. Iversen, J. K. Møller, J. M. Morales, H. Madsen. Inhomogeneous Markov Models for Describing Driving Patterns. *submitted, 2015.*

[B] E. B. Iversen, J. M. Morales, H. Madsen. Optimal Charging of an Electric Vehicle using a Markov Decision Process. *Applied Energy, 123, pp. 1–12, 2014.*

[C] E. B. Iversen, J. K. Møller, J. M. Morales, H. Madsen. Probabilistic Forecasts of Solar Irradiance using Stochastic Differential Equations. *Environmetrics, 25, pp. 152–164, 2014.*

[D] E. B. Iversen, J. K. Møller, J. M. Morales, H. Madsen. Short-term Probabilistic Forecasting of Wind Speed using Stochastic Differential Equations. *International Journal of Forecasting, in print, 2015.*

[E] E. B. Iversen, J. K. Møller, J. M. Morales, Pierre-Julien Trombe, H. Madsen. Leveraging Stochastic Differential Equations for Probabilistic Forecasting of Wind Power using a Dynamic Power Curve. *Submitted, 2015.*

[F] E. B. Iversen, R. Juhl, J. K. Møller, J. M. Morales, J. Kleissl, H. Madsen. Spatio-Temporal Forecasting by Coupled Stochastic Differential Equations: Applications to Solar Power. *Submitted, 2015.*

[G] N. Juul, G. Pantuso, E. B. Iversen, T. K. Boomsma. Strategies for Charging Electric Vehicles in the Electricity Market. *Submitted, 2015.*

Presentations

In addition to the papers listed before, the following talks were presented at conferences during the project period.

[H] J. K. Møller, E. B. Iversen, H. Madsen. Probabilistic wind power forecast by stochastic differential equation models. *The 23rd Annual TIES Conference, June 2013, Anchorage, Alaska, USA.*

[I] E. B. Iversen, J. M. Morales, H. Madsen. Optimal Charging of an Electric Vehicle Considering User Preferences . *Danish Smart Grid Network, March 2013, Copenhagen, Denmark.*

[J] E. B. Iversen, J. M. Morales, H. Madsen. Driving patterns and Optimal Charging of an Electric Vehicle. *Smart-Grid Workshop, August 2012, Lund, Sweden.*

Contents

Part I

Summary Report

Chapter 1

Introduction

1.1 Motivation

Renewable energies and in particular wind and solar power have seen rapid growth in the last decade. This growth rate is expected to continue in the immediate future as projects under way come online and in the more distant future renewable is expected to see a higher penetration ([Age14]). The growth is driven by several factors such as a growing energy demand, increasing prices of fossil fuels, a need to reduce carbon emissions and security of energy supply. These issues are not expected to relent in the coming decades. The demand for energy world wide is expected to keep growing even with increased energy efficiency. Fuel prices are expected to increase as a function of increased demand coupled with diminishing supply. Global warming calls for alternatives to using fossil fuels and for thinking of alternative energy sources. On top of this it is a national security issue for most nations to secure a stable and affordable energy supply for their populations. All of this makes renewable energy sources attractive and as such their penetration is expected to grow in the coming years.

The penetration of renewable energy in the electrical grid is already substantial in many European countries. In Germany net-generation from renewable energy sources constitute almost 30 % of the electricity consumption increasing from 6 % in year 2000 ([fra]). In the first half year of 2014, 41 % of the Danish electricity consumption

was supplied from wind power generation with and ambition to reach 100 % in 2035 ([Hov14]). This calls for significant changes to the existing energy system.

Electricity is a commodity that has the very particular feature that it cannot be stored - it has to be consumed at the same time as it is generated. Thus for power systems management the main issue to solve has been to balance out consumption with generation at all points in the electrical grid at all times. Conventionally this is carried out by having stored energy such as chemical energy in terms of fossil fuels, potential energy in terms of hydro power or nuclear energy as uranium that are converted to electrical energy at the appropriate time such that supply follows the demand. Here power stations are dispatched to meet the demand of the consumers.

Renewables such as wind or solar power are in this way fundamentally different from conventional sources. Wind and solar power cannot be dispatched at the power producers discretion - at night there is no sun and if the wind does not blow there is no wind power. Due to this specific intermittent and variable nature of wind and solar power it is often referred to as stochastic. With a small penetration of stochastic generation in the electrical grid the uncertainties have only minor effects on grid stability. Thus it may be feasible to consider the supply uncertainties to be part of the demand side uncertainty, which is already modeled in most power system models. Modeling the uncertain generation in this way, increases demand side uncertainty and thus increases the demand for generation reserves. This may be viable for low penetrations of renewable generation. However for large scale integration, this is both economically expensive and can potentially offset the emissions benefits obtained from using renewable energy. Thus for large scale integration of renewables, increased demand side uncertainty is not viable.

Instead of assuming large uncertainties on the demand side, a more sensible approach is to directly consider the electrical supply as stochastic. Some uncertainty is already considered on the supply side such as failures of large integrated generation or failure of the grid in terms of interconnectors. This leads us to possible approaches to solve the issued of supply side uncertainty:

- *Upgrading the grid* by providing better interconnectors allowing power to flow from areas with surplus production to areas with surplus demand. Though this promises to reduce the instability in the grid it is unlikely to entirely mitigate the large scale introduction of supply side uncertainty will have on grid stability.
- *Energy storage* which may consist of batteries, pumped hydro or other options that present a two way conversion between electrical and some other energy form. For energy storage to be viable it needs to both be cheap and relatively free of losses in the conversion.
- *Demand response* consists in shifting the time of demand for electricity by the

consumers. Electric vehicles, electrical building heating, heating of water tanks could act as flexible consumption helping to match the electricity demand to the supply.

- *Forcasting of uncertain generation* as input to decision making problems in the power grid has a substantial history. Forecasts allow grid operators to make decisions ahead of actual problems alleviating issues before they arise. Today forecasts of wind and solar power generation are widely used by many electrical utilities with a track record of reducing operating costs and improving grid stability.

Forecasts of the uncertain generation from solar and wind power do not constitute the solution to the issue of uncertain and intermittent generation. However, as these forecasts often serve as input information to grid operation, they are vital for an efficient use of storage and they allow the electrical demand to respond to the varying power production. Thus forecasts constitute a necessary and cost efficient part of the transition to a power system with a high penetration of uncertain generation. Also it becomes clear that all forecasts are not equal. Forecasts of high quality thus provide decision makers with more information and thus allows for better decisions. We shall return to the concept of forecast quality.

1.2 Thesis Objective and Outline

The objective of this thesis is to advance the stat-of-the-art in terms of forecasting for renewable energy systems by improving the quality of forecasts and show how to apply these improved forecasts.

Forecasting for energy systems ranges from forecasting the power output from a specific wind farm at some point in the future to forecasting the distribution of power production for individual wind farms over an area for multiple time horizons. On the demand side forecasting ranges from predicting the total power consumption for a specific area for some point in time to predicting the probability of a certain amount of power consumption by a single consumer over some time period. With forecasts being so varied in nature the challenge is how to best use these forecasts and particularly what is the most appropriate forecast for a specific application.

The purpose of this research is to develop models that provide the necessary input to decision makers with the main emphasis being on providing forecasts. Different types of forecasts are of interest: point forecasts, univariate density forecasts, density forecasts on multiple horizons finally considering spatio-temporal predictive densities considering the interdependence between different locations and times.

We first consider a closed case study that is a single electric vehicle. In Paper A the issue of when a single electric vehicle is in use is considered. Subsequently a data driven model for the use of this vehicle is developed based on in-homogeneous Markov chains. This model is particular in the sense that it does not provide a single output forecasts but a probability of being in a given state at some point in the future.

Building on top of the model for the use of a single vehicle in Paper A, Paper B devises a strategy for charging the vehicle in such a way that the vehicle utilizes varying electricity prices and is charged at the minimal costs while still satisfying the users driving preferences. This is done by assigning a penalty, that the end user can decide, for the vehicle not being able to comply with the users driving needs. The charging strategy builds on stochastic dynamic programming and a Markov decision process.

Paper G goes on to consider the economic benefits of different charging algorithms for a fleet of electric vehicles considering a stochastic power price. Here we find that increasing the information available, decreases the expected cost.

Going into forecast models for the supply side of the electrical grid, Paper C introduces a novel methodology for forecasting using stochastic differential equations applied to solar irradiance forecasting. A probabilistic forecast model is developed exploiting the special features of continuous time modeling to provide predictive densities for solar irradiance at a single location.

The basic methodology introduced for forecasting introduced in Paper C is in Paper D applied to forecast wind speeds and further extended to better exploit external input such as numerical weather predictions into the forecast.

Paper E builds upon Paper D and introduces a dynamic power curve transforming predicted wind speed into predicted power for a wind farm in Denmark. In this paper further emphasis is put on multi-horizon forecasts and on how a proper description of the predictive density paves the way for proper multi-horizon forecasts.

Spatio-temporal forecasts present a challenge as they are both computationally and parameter intensive. In Paper F the specific advantages of modeling in continuous time are applied to spatio-temporal models. Here a stochastic differential equation framework is used for modeling which in turn can be interpreted as a discretized version of stochastic partial differential equation model. Thus this yields a model that is both computationally efficient and has a low dimensional parameter space.

The remainder of thesis is structured as follows: In Chapter 2 we go into why forecasts for renewable energy production presents a challenge. Also we go into detail with what the quality of a forecast means and how this the quality is related to how the forecast is to be used. This leads into *classical* approaches to modeling and forecasting for

uncertain systems, which is covered in Chapter 3. We go on to address the specifics of the approaches taken in the papers covered in the thesis. In Chapter 4 we take a look at the different types of forecasts products and how they are used in practice in relation to energy systems. Chapter 5 concludes the thesis outlining the main contributions and gives perspectives for future research within this topic.

Chapter 2

Challenges of Forecasting in Energy Systems

In Chapter 1 we briefly mentioned that renewable energy generation introduce uncertain generation into the power grid. This chapter discusses the importance of forecasting in energy systems and the challenges associated with this forecasting.

Forecasts of energy consumption and generation are essential for an efficient integration of renewables into electricity markets operations, as the markets should be cleared in advance and market participants must make decisions even before that. This holds true for all types of electricity markets spanning from real-time markets to futures markets to day-ahead markets ([mor]). Forecasts for conventional generation is, in comparison to forecasts for renewable generation, rather straight forward as the future electricity generation can be controlled directly, except for unit failures. Thus for conventional generation forecasts consist of potential schedules, that are translated into supply in the market. As for renewable generation meteorological phenomenons have a hand in determining the future schedule of the power plants. Wind turbines only produce power when the wind blows, solar power plants produce only power during the day and during the day the amount of power varies due to cloud cover and sun height. While hydro energy has an inherent storage capability it may not produce power if the reservoir is empty or it maybe be constrained due to flood control. The inherent variability and non-storability of wind and solar power generation (as opposed to hydro) increases the need for appropriate forecasts for this

type of generation.

Some argue that forecasts are mainly there to comfort decision makers (these decision makers being the market and network operators), power producers and to some extent end users, while they are not used or not used optimally in day-to-day operations. However, using appropriate forecasts for well defined decision making problems can improve the quality of the decisions immensely ([mor]). Thus, it makes sense to consider not forecasting and decision making as two distinct problems, but as two parts of one integrated problem that needs to be solved. In this way forecasting models should not be developed independently of the decision making problem as likewise solutions for decision making problems should also take into account the most appropriate input information to yield the solution of the highest quality.

Challenges from forecasting renewable energy production stems from the inherent nature of the weather. The global weather system is best described as a chaotic system. In chaotic systems any uncertainty in the initial conditions will produce rapidly escalating and compounding errors in the prediction of the systems future behavior. To make predictions of long term behavior without errors, the initial conditions must be know in their entirety and with infinite accuracy. This is of course impossible. Further more the forecast model must have no model errors or numerical approximations. This does not hold in general for systems with any significant complexity. The forecast errors from the weather system may even be exacerbated by the conversion from wind to wind power and solar irradiance to solar power by the non-linearities in this conversion. This leads us to the conclusion that all forecasts are *wrong*, in the sense that they do not predict the future with certainty. However, that a system is chaotic and that we can never hope to predict the future exactly does not mean that we can predict nothing at all. This brings up the question: what constitutes a good forecast and what makes it better than another forecast?

To rank forecasts among each other, we introduce the terms quality and value of a forecast. We define the quality of a forecast as some statistical predefined magnitude for which we get a single output value by which we can rank the forecasts. Notice that the statistical quality of a forecast does not consider the application of the forecast. Defining quality in this way, there is no unique way of ranking forecasts among each other, this depends on the actual choice of criteria. Similarly we can rank forecasts in terms of the value that they provide to the end user. Ranking forecasts in this way in turn depends on the application and on what the end user is able to do with said forecast. With this value approach, the problem of forecasting can not be seen as distinct from the application and the forecasting and optimization problem should thus be seen as two parts of the same overall challenge. In the Example 2.1 we provide an example of optimal bidding for a unit with uncertain generation.

Example 2.1 To illustrate how the appropriate forecasts depends on the application consider the stylized case of a wind power producer selling her production in the market ahead of delivery. The wind power producer is tasked with maximizing expected profits. Also the wind power producer can only bid a production quantity into the market and thus is a price taker. Furthermore if the amount that the wind power producer bids does not match the actual generation she is penalized in relation to the size of the error. Thus the wind power producers problem is to bid the amount of power generation, $\tilde{x}$, that maximizes the expected profits, i.e.

$$\max_{\tilde{x}} \left\{ \mathbb{E}_X \left[\pi(\tilde{x}) \right] \right\}, \tag{2.1}$$

where $\pi(\cdot)$ is the profit and X is the uncertain future power generation. Now suppose that the power producer is faced with one penalty for up-regulation p_{up} and another for down regulation p_{dw}. We thus get the profit:

$$\pi(\tilde{x}) = p\tilde{x} - p_{\text{up}}(\tilde{x} - X)^+ + p_{\text{dw}}(\tilde{x} - X)^-, \tag{2.2}$$

where $(\cdot)^-$ and $(\cdot)^+$ is the negative part and the positive part respectively. Now consider:

$$\begin{aligned} \mathbb{E}_X \left[\pi(\tilde{x}) \right] &= p\tilde{x} - p_{\text{up}} \mathbb{E}\left[(\tilde{x} - X)^+ \right] + p_{\text{dw}} \mathbb{E}\left[(\tilde{x} - X)^- \right] && (2.3) \\ &= p\tilde{x} - p_{\text{up}} \int_0^{\tilde{x}} (\tilde{x} - s) f(s) ds + p_{\text{dw}} \int_{\tilde{x}}^{\bar{x}} (s - \tilde{x}) f(s) ds, && (2.4) \end{aligned}$$

where $\bar{x}$ is the rated capacity of the wind farm. Set $\frac{\partial}{\partial \tilde{x}} \mathbb{E}_X(\pi(\tilde{x})) = 0$ to find the optimum, which yields:

$$\begin{aligned} 0 &= p - p_{\text{up}} \int_0^{\tilde{x}} f(s) ds - p_{\text{dw}} \int_{\tilde{x}}^{\bar{x}} f(s) ds && (2.5) \\ &= p - p_{\text{up}} F(\tilde{x}) - p_{\text{dw}}(1 - F(\tilde{x})). && (2.6) \end{aligned}$$

Suppose now that the costs paid are symmetric. This can be done by setting $p_{\text{up}} = p + c$ and $p_{\text{dw}} = p - c$. Inserting this into the above equation yields:

$$\begin{aligned} 0 &= p - (p + c)F(\tilde{x}) - (p - c)(1 - F(\tilde{x})) \Rightarrow && (2.7) \\ \tilde{x} &= F^{-1}(0.5). && (2.8) \end{aligned}$$

Thus the optimal bid for maximizing the expected profit is bidding the median of the predictive distribution of the wind power production. The median is a classical forecasts output also corresponding to the mean if the distribution is symmetric. However if the penalty scheme is changed to be non-symmetric, the optimal bid changes. The above calculations can be repeated $p_{\text{up}} = p + c_{\text{up}}$ and $p_{\text{dw}} = p - c_{\text{dw}}$ to find the optimal bid being $\tilde{x} = F^{-1}\left(\frac{c_{\text{up}}}{c_{\text{up}} + c_{\text{dw}}} \right)$. Thus, the optimal bid turns out to be a quantile in the predictive distribution of the power generation. Hence, to

bid optimally in this very simple market setup a forecast that produces quantiles is required.

As illustrated in Example 2.1 forecasts and their application should be considered in unison. This holds not only for the above example but for many forecast application. The forecast considered should contain the appropriate information. Through this thesis we will see several examples that underlines this point: That forecasts and applications should be developed in unison.

The challenges for forecasters are not limited to naively supply forecast users with forecasts that minimize some criteria such as squared deviations. While forecast that solve this problem may be appropriate for some applications, this does not hold true in general. The real challenge for forecasters lie in understanding the forecast users problem and to supply them with the forecast that best helps him solve their problem. This means that forecasters, depending on the application, should be able to, but not limited to, provide forecast products such as point forecasts, predictive densities and scenarios.

Chapter 3

Mathematical Tools for Forecasting

Forecasting is the process of making statements about outcomes that have not been observed yet. As knowledge about the future (or the unobserved) is valuable in many fields, forecasting has seen many and diverse applications. Examples of forecast applications range from finance where models are used to forecast stock returns ([CGS08]), petroleum futures volatility ([Sad06]) to default risk ([Ati01]). To biological applications where forecasts are used to estimating future fish stocks ([PCD+01]), predicting epidemics ([HBG04]) and to predicting invasive species ([LD03]).

A forecast can be obtained in many ways ranging with two general categories: qualitative methods, basing their forecast on the opinion and judgment of experts, an quantitative methods, using data and mathematical models to make predictions about the future. The forecasts that we consider in this thesis are outputs of mathematical models or forecast models. As such forecast models can be classified further in the following categories: Linear vs. nonlinear, static vs. dynamic, explicit vs implicit, discrete vs. continuous, deterministic vs. probabilistic and deductive vs. inductive. While most mathematical models can be classified in a straight forward way, describing all different combinations in detail becomes a monumental task. In this thesis we adopt the distinction between deductive models and inductive models. This classification can also be expressed as models that a primarily driven by a physical understanding of the system at hand versus those that are driven by data. We adopt

this distinction as the two communities of forecasters are also distinct from each other.

The main contribution of this thesis lies in the realm of *gray box* models, that are a crossover between physical models and data driven models. As such they model relationships obtained from the physics of the system at hand to improve the data driven model structure. This allows for an intuitive modeling approach which can yield models that outperform both physical and data-driven models. We now go into some of the forecast methods applied in energy systems operation and in particular for energy systems with a large stochastic component.

3.1 Physical Models

Models based on physics are often referred to as white or glass box model, as the inner workings of the models are clearly understood. Physical models for energy systems applications are typically used for grid stability forecasting or for point predictions of renewable generation. Producing forecasts of renewable energy generation from a physical model is the result a detailed physical understanding of meteorological phenomenona and how this translates into changes in the power produced. Forecast models for grid stability, based on a physical understanding of the power system, focuses on predicting the reserves need, should a rare event happen. Thus physical models for power systems operation take into account the entire system. Here the single point failure that will most affect the power system is identified. This is then typically assumed as the worst possible scenario and used for setting the reserve capacity in the power system. In Example 3.1 we provide an example of a physical approach to forecasting generated power from a wind farm in northern Denmark.

Example 3.1 To illustrate how a physical model can produce a forecast of renewable energy generation a short example is presented here of how to obtain such a forecast. Suppose we want to obtain a forecasts for power generation for the Klim wind farm located in the northern part of Denmark consisting of 35 Vestas V44/600 wind turbines each with a nominal generation capacity of 600 kW.

First atmospheric physics are used to model the meteorological system through an atmospheric model. Various examples of such physical models are, among others, HIRLAM ([K96]), ALADIN ([BHBG95]), WRF ([SKD+05]) and HH5 ([GDS+94]). In these models the atmosphere is described as a fluid. Hence, the models base them selves on fluid dynamics and thermodynamics and are based on the Navier-Stokes equations and the laws of thermodynamics. Data is entered into the model through a process called initialization where meteorological observations are fed into

the model. The main inputs are from weather satellites and weather balloons with a data resolution down to 1 km. The atmospheric model is then run to produce meteorological information for future times given the location and altitude. These include the air density, pressure, temperature fields and, most interestingly for our application, air velocity (wind) fields. The equations used are nonlinear partial differential equations, where no exact solutions exist. Thus numerical methods are used to approximate these solutions.

With a forecast wind speed at hand, we can now proceed to translate this predicted wind speed into power. The available power in a constant stream of air with the same cross section can be shown to be given by:

$$E = \frac{1}{2}\rho v^3 \frac{\pi d^2}{4}, \tag{3.1}$$

where v is the wind speed, ρ is the air density, and d is the cross section diameter. Next the physics of the wind turbine should be considered. There are physical limitations on the rotor speed, friction, operational settings and other physical constraints related to the wind turbine. All this is captured in the so called power curve. The power curve for a Vestas V44/600 is shown in Figure 3.1. Notice the cut-in wind speed at 5 m/s and the cut-out wind speed at 20 m/s. Also notice that the power curve is highly nonlinear.

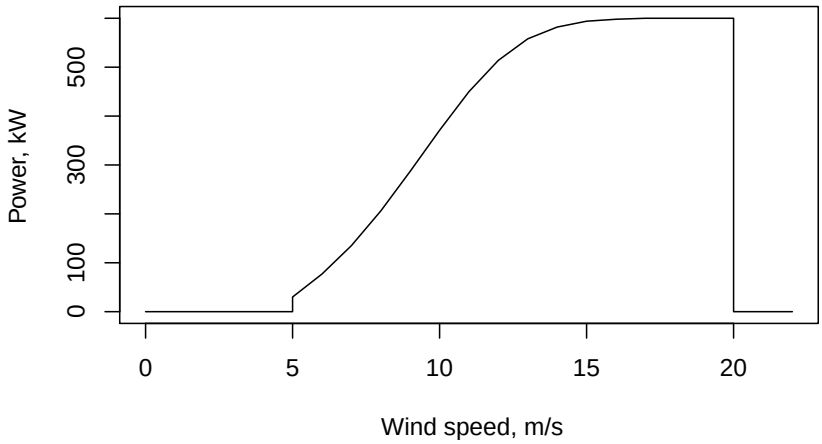

Figure 3.1: *The power curve for a Vestas V44/600 wind turbine.*

Following this, predicting the output power of the Klim wind farm should be straight forward. Simply translate the predicted wind speed into wind power by the power curve and multiply by the number of wind turbines.

As previously mentioned, solving the partial differential equations needed for a physical model introduces numerical approximations and thus errors. Thus we can never expect to predict the future power output exactly. These errors can be systematically reduced in different ways such as combining numerical weather predictions from different physical models to increase the performance of the prediction system, as done in [NNM+07], or used in conjunction with the Kalman filter to remove systematic errors in the NWP forecasts, as proposed in [LGS+08]. This leads into the topic of data driven models.

3.2 Data Driven Models

Data driven models, as opposed to physical models, rely on the available data as opposed to physical reasoning. These models are also often referred to as black box models. The name comes from the underlying dynamics being hidden in a black box and only the inputs and outputs are known. This is opposed to white box models where there is a clear physical relationships that governs the input to output relation. Forecasting comes into its own when the challenge is to try and predict the outcome of systems with a level of complexity that makes physical models impractical or impossible.

In energy systems applications many types of black box model have been applied. These range from predicting power or heat load, predicting power prices, forecasting renewable power generation to modeling consumer response to power price changes. While many approaches exist for this type of forecasting the different approaches may overlap substantially. Some main categories include:

- *Time series methods* rely on using historical data to make predictions about the future. Classical approaches include Box-Jenkins models (Auto-Regressive, Auto-Regressive-Moving-Average, Auto-Regressive-Integrated-Moving-Average among others), Kalman filter models and different types of smoothing approaches. Box-Jenkins type models have been used to forecast wind power generation ([PM12]), to forecast power prices in the Nordpool area ([Kri12]) and to forecast solar power generation ([Rei09]). An example of a Kalman filter approach to wind power forecasting can be found in [PPP13].

- *Causal forecasting methods* identify underlying factors that might influence the variable being forecast and use these factors to predict the future outcome of the variable in question. This can be done by regression type models which have been applied to power load forecasting in [CCVO98]. Auto-regressive models with external inputs are sometimes seen as causal methods and some times as

time-series methods and have been applied to forecast power generation from a photo-voltaic system in [LSS14].

- *Artificial intelligence methods* which include artificial neural networks and support vector machines use advanced algorithms to estimate or approximate functions that depend on a large number of inputs and are generally unknown. This takes a large part of the modeling out the hands of the modeler and instead uses designed algorithms. For energy systems application data mining and fuzzy logic has been used to forecast wind power generation ([KZS09], [DATD04], [PN06]).

In Example 3.2 we introduce a data driven approach to forecasting the power output of the same wind farm as mentioned in Example 3.1. This data driven approach relies only on data and not a physical understanding of the system.

Example 3.2 Referring to the problem presented in Example 3.1, consider a black box modeling approach for generating these same forecasts. We now present an auto regressive external input model with generalized auto regressive conditional heteroskedasticity (or ARX-GARCH for short) used as a benchmark model in Paper D. The model takes the following form:

$$Y_k = \psi_0 + \sum_{i=1}^{q} \psi_i Y_{k-i} + \phi p_{t_k} + \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}(0, \sigma_k^2) \tag{3.2}$$

$$\sigma_k^2 = \alpha_0 + \sum_{i=1}^{\tilde{q}} \alpha_i \sigma_{k-i}^2 + \sum_{i=1}^{\bar{q}} \beta_i \tilde{\epsilon}_{k-i} \qquad \tilde{\epsilon}_k \sim \mathcal{N}(0, \tilde{\sigma}^2). \tag{3.3}$$

Here Y_k is the power production at observation k, p_{t_k} is an external input that is the predicted power obtained from applying a local regression model on the predicted wind speeds to predict the power generation. $\theta_i, \phi, \alpha_i, \beta_i$ and σ are parameters and $q, \tilde{q}$ and $\bar{q}$ determine the number of lags in the model. σ_k^2 describes the time varying variance of the wind power process.

In this model the parameter can be found by maximizing the likelihood under the model. This model has the advantage over the model in Example 3.1 that it is calibrated to data. Also it can provide predictive densities, which the model presented in Example 3.1 cannot.

A somewhat obvious question may arise as to why not try and combine the physical models with the data driven models? This is one of the major tasks undertaken in this thesis an we go through some of the results in the following section.

3.3 Gray Box Models

Gray box models obtain their name from their nature of being a mixture of black and white box models. Gray box models attempt to describe the system in question partly by physics and partly by data. Gray box models can assume many forms depending on the problem at hand. The need to integrate data into the model has the side-effect that they can produce predictive densities, that is the output extends beyond a single point or time-trajectory to predicting a distribution of the outcome. Another feature is that gray box models typically model dynamics continuous in time. This is due to the fact that the real world does not have a minimal time increment and as such these models are similar to physical models and dissimilar from the typical data driven model. An advantage of continuous time models is that they should (in theory) be able to be translated across different time discretizations. The typical gray box model describes some underlying state of the system at hand and is referred to as a state space model.

Gray box models typically model dynamics in continuous time, similar to physical models, but have observations in discrete time, characteristic of most data driven models as data can typically not be observed continuously. This modeling in continuous time allows the models to be formulated on a more intuitive form and to reduce the number of parameters, while still conforming to the discrete time observations. The drawbacks include the need for somewhat advanced mathematical methods and also the complicated mathematical structure of the models can make parameter estimation difficult.

3.3.1 Discrete State Space

Discrete state space models are characterized by having a finite number of outcomes. This can be exemplified by a switch being on or off, a power plant being in production, in stand-by or shut-down or the occupancy of a room, where the states may describe the number of occupants in the room. As the discrete state space models only rely on few assumptions on the underlying data, they may be seen a dark gray box model, in the sense that they are closer to the black box model than the white box model. In Example 3.3.1 we present an application of this type of model from Paper A, where a discrete state space model is applied to modeling driving patterns of a single vehicle.

A continuous time Markov chain, X_t, is defined by a finite or countable state space $\mathcal{S}$ and a transition rate matrix $\mathbf{Q}$ with dimensions equal to that of the state space and an initial distribution. For $i \neq j$, the elements of $\mathbf{Q}$, q_{ij} are non negative and describe the rate of transition from state i to state j. The element q_{ii} is defined such that the sum of each row is zero. For infinitesimal increments in time, h, we have

that:

$$P\left(X_{t+h} = j | X_t = i\right) = q_{ij}h + o(h) \tag{3.4}$$

where o is the little-o notation. This can be interpreted as the chain can only perform one transition in infinitesimal time. This leads to some nice structural abilities for continuous time Markov chains as we shall see in example 3.3.1, where a continuous time Markov chain is applied to model vehicle driving patterns (the model is developed in Paper A).

Example 3.3.1 Consider the problem of modeling a single vehicle where we want to model the vehicle being parked or in used/driving. Here, clearly the model should have one of two possible outcomes at each time, namely that the vehicle is driving or not-driving. A Markov chain is thus the obvious choice for this model.

While the real world has no "time step", it is in general not possible to observe anything continuous in time, that is data is collected in discrete packets. Thus there is a need for a translation between the continuous time nature of the world and the discrete time observations. The transition intensities of a Markov chain with two states are given as the transition matrix $\mathbf{Q}(u)$ given by:

$$\mathbf{Q}(u) = \begin{pmatrix} -q_{11}(u) & q_{12}(u) \\ q_{21}(u) & -q_{22}(u) \end{pmatrix} = \begin{pmatrix} -q_{12}(u) & q_{12}(u) \\ q_{21}(u) & -q_{21}(u) \end{pmatrix}, \tag{3.5}$$

where $q_{jk}(u)$ is the transition intensity between state j and state k at time u. The probability of transitions between the two states are then given by the transition probability matrix $\mathbf{P}(t)$ contains the probability of moving between the different states between t and $t+1$, i.e. in one time step. Suppose that $\mathbf{Q}(u)$ is constant over this time step, which may be reasonable if the time step is small enough. Then the one minute transition probabilities are given by:

$$\mathbf{P}(t) = \begin{pmatrix} p_{11}(t) & p_{12}(t) \\ p_{21}(t) & p_{22}(t) \end{pmatrix} = \begin{pmatrix} 1 - p_{12}(t) & p_{12}(t) \\ p_{21}(t) & 1 - p_{21}(t) \end{pmatrix} = e^{\mathbf{Q}(t)}. \tag{3.6}$$

In the formulation with two states there is no immediate benefit from using a continuous time model as they produce the same output and have the same number of parameters. However, a continuous time Markov chain model will allow for a parameter reduction if certain structures are present. Furthermore, identifying such structures will make the model more theoretically tractable. One of these requirements is that there are more than two states in the model.

In Figure 3.2 the estimated transition probabilities are shown and obtained after a parameter reduction using B-splines to fit the time varying transition probabilities.

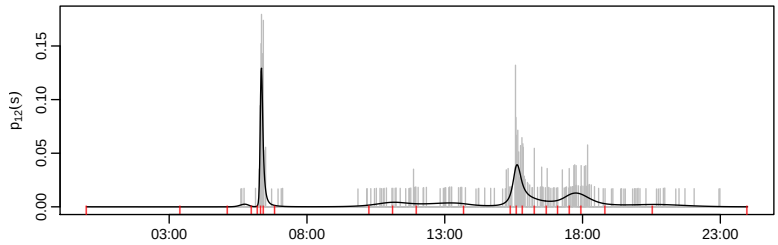

Figure 3.2: *$\widehat{p}_{1\cdot}(s)$ based on the B-splines and the logistic regression, plotted as the black line over the estimates $\widehat{p}_{1\cdot}(s)$ in gray. The red bars indicate the knot positioning for the B-splines.*

As a simple illustration of such a model, consider the case where there are four states, i.e. $N = 4$. State 1 corresponds to the vehicle being parked at home. State 2 corresponds to the vehicle being on a trip that started from home. State 3 corresponds to the vehicle being parked somewhere else. State 4 corresponds to the vehicle starting a trip from somewhere else than at home. The parameter reduction is obtained if it is assumed that the vehicle cannot switch directly from being parked at home to being parked somewhere else, that is a transition from states 1 to 3 cannot occur. Also it would be reasonable to assume that the vehicle does not drive from home to return to home, without an intermediate stop. Under these assumptions, the matrix of transition intensities becomes:

$$\mathbf{Q}(u) = \begin{pmatrix} -q_{12}(u) & q_{12}(u) & 0 & 0 \\ 0 & -q_{23}(u) & q_{23}(u) & 0 \\ 0 & 0 & -q_{34}(u) & q_{34}(u) \\ q_{41}(u) & 0 & q_{43}(u) & -(q_{43}(u) + q_{41}(u)) \end{pmatrix}. \tag{3.7}$$

Thus we can use an understanding of the physics in the system to reduce the number of parameters in the model and to obtain better estimates. Here the number of parameters are reduced from 12 to 5.

Models for driving patterns are of interest within energy applications due to the many electric vehicles being introduced in recent years. Managing the charging of electric vehicles have seen interest from the energy community, both with grid stability in mind and to absorb excess generation from renewable source. Hence, understanding how vehicles are used becomes of fundamental importance for doing smart charging as not to inconvenience the end user as we will see in Example 4.4.

3.3.2 Continuous State Space

Many variables are not limited to a finite set of possible values as for discrete state space models. These models are typically continuous state space model, where the variable in question lives on a continuous sample space that is some subset of $\mathbb{R}^N$. Continuous sample space models have many applications in energy systems ranging from forecasting power generation from renewables, price forecasts, load forecasting and in general to dynamical problems relating to time series. The contribution in the thesis relating to this field can be found in papers C, D, E, and F.

The physical models described in Section 3.1 are typically described by differential equations and as such this differential equations provide a starting point for continuous time gray box models. In the ordinary differential equation setting, the evolution in time of the state variable, X_t, is given by the deterministic system equation

$$\frac{dX_t}{dt} = f(X_t, t), \tag{3.8}$$

where $t \in \mathbb{R}$ and $f(\cdot) \in \mathbb{R}^n$. As previously mentioned, complex systems such as weather systems are subject to random perturbations of the input or processes that are not specified in the model description and numerical errors. Thus the deterministic model will never predict the future output precisely. This suggests introducing a stochastic component in the state evolution to capture these perturbations or to capture model deficiencies. This can be done by formulating the state evolution as a stochastic differential equation (SDE), as done in [Øks10]. Thus, we can formulate the time evolution of the state of the process in the form:

$$\frac{dX_t}{dt} = f(X_t, t) + g(X_t, t)W_t, \tag{3.9}$$

where $W_t \in \mathbb{R}^m$ is an m-dimensional standard Wiener process and $g(\cdot) \in \mathbb{R}^{n \times m}$ is a matrix function [Øks10]. Multiplying by dt on both sides of (3.9) we get the standard SDE formulation:

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t. \tag{3.10}$$

While this form is the most common for SDEs, it is not well defined, as the derivative of W_t, $\frac{dW_t}{dt}$, does not exist. Instead, it should be interpreted as an informal way of writing the integral equation:

$$X_t = X_0 + \int_0^t f(X_s, s)ds + \int_0^t g(X_s, s,)dW_s. \tag{3.11}$$

In Equation (3.11), the behavior of the continuous time stochastic process X_t is expressed as the sum of an initial stochastic variable, an ordinary Lebesgue integral, and an Itō integral.

In a deterministic ordinary differential equation setting, the solution would be a single point for each future time t. In the SDE setting, in contrast, the solution is the probability density of X_t for any state, x, and any future time, t. For an Itō process given by the stochastic differential equation defined in (3.10) with drift $f(X_t, t)$ and diffusion coefficient $g(X_t, t) = \sqrt{2D(X_t, t)}$, the probability density $j(x, t)$ in the state x at time t of the random variable X_t is given as the solution to the partial differential equation known as the Fokker-Planck equation [Bjö09]:

$$\frac{\partial}{\partial t} j(x,t) = -\frac{\partial}{\partial x}\left[f(x,t)j(x,t)\right] + \frac{\partial^2}{\partial x^2}\left[D(x,t)j(x,t)\right]. \tag{3.12}$$

Hence, given a specific SDE, we can find the density at any future time by solving a partial differential equation.

This type of model for forecasting was used in and developed in Paper C, further developed and applied in Paper D, and in Paper E. In Example 3.3.2 we introduce a gray box approach to forecasting wind power production.

Example 3.3.2 We revisit the problem presented in Example 3.1 and in Example 3.2 and propose a gray box model modeling and forecasting this same wind farm. The forecasts are developed in Paper E. The model takes the following form:

$$\begin{aligned}
dX_t &= \left(\left(1 - e^{-X_t}\right)\left(\rho_x \dot{p}_t + R\right) + \theta_x(p_t\mu_x - X_t)\right)dt + \sigma_x X_t^{0.5} dW_{x,t} && (3.13)\\
dR_t &= -\theta_r R_t dt + \sigma_r dW_{r,t} && (3.14)\\
Y_{1,k} &= X_{t_k} + \epsilon_{1,k} && (3.15)\\
dQ_t &= (S_t - \theta_q Q_t)dt + \sigma_q dW_{q,t} && (3.16)\\
dS_t &= -\theta_s S_t dt + \sigma_s dW_{s,t} && (3.17)\\
Y_{2,k} &= \left(0.5 + 0.5\tanh(5(X_{t_k} - \gamma_1))\right)\left(0.5 - 0.5\tanh(\gamma_2(X_{t_k} - \gamma_3))\right) && (3.18)\\
&\quad \frac{\zeta_3}{1 + e^{-\zeta_1(X_{t_k} - \zeta_2 + Q_{t_k})}} + \epsilon_{2,k}. && (3.19)
\end{aligned}$$

In the above equations (3.13)-(3.15) model the wind speed dynamics and equations (3.16)-(3.19) model a dynamic power curve translating these wind speeds into power production. In the model defined in equations (3.13)-(3.19) X_t, R_t, Q_t, S_t, $Y_{1,k}, Y_{2,k}, \epsilon_{1,k}$, and $\epsilon_{2,k}$ are stochastic variables, with $Y_{1,k}$ being the observed wind speed and $Y_{2,k}$ being the observed power. p_t is the numerical weather prediction and $\dot{p}_t$ is the increment in the numerical weather prediction between time $t-1$ and t. $\rho_x, \theta_x, \mu_x, \sigma_x, \theta_r, \sigma_r, \theta_q, \sigma_q, \theta_s, \sigma_q, \gamma_1, \gamma_2, \gamma_3, \zeta_1, \zeta_2, \zeta_3$ are parameters in the model that are estimated. For a further explaination of the model see Paper E.

In the model presented in equations (3.13)-(3.19), the diffusion term in equation (3.13) drops to zero, when X_t approaches zero. Therefore, with X_t close to zero, the

process is dominated by the drift term. Furthermore, the drift term always has drift away from zero. Together this results in a process that lives on $\mathbb{R}^+$, and has higher numerical variability for higher numerical values, which is exactly like the wind speed actually behaves. Further more, equation (3.15) describes the power curve, that is highly non-linear and also confines the normalized power to the domain $[0, 1]$.

Among the main contributions from the model presented here as opposed to those in Example 3.1 and Example 3.2 is that the predictive density is varying depending on the predicted output power. This is due to the highly nonlinear dynamic power curve and the probabilistic model for wind speeds. This along with the model for the wind speed dynamics help capture the skewed and bounded nature of the predictive densities for wind power generation. An example of this can be seen in Figure 3.3, where predictive densities for normalized power production for forecast horizons of 1 and 24 hours ahead are shown.

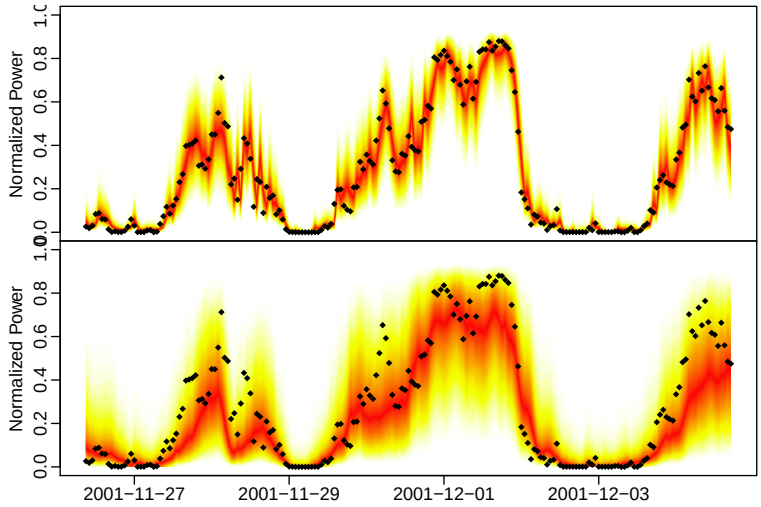

Figure 3.3: *1-hour (top) and 24-hour (bottom) ahead predictive density of model (3.13)–(3.19), with warmer colors indicating a higher probability of seeing this realization. The densities are approximated by Monte Carlo simulations.*

An added benefit of using this type of model is that a more accurate description of the system noise allows for predicting for multiple horizons from the same model without the loss in performance seen in black box models, as we show in Paper E.

3.3.3 Spatio-Temporal Forecasting

Space time problems provide many challenges related to the large number of outputs and the large number of inputs. This makes them computationally intensive and particularly tricky to estimate. For this reason the typical approach takes it offset in physical models letting the dynamics be governed by laws of physics. This has the obvious drawback of neglecting the model errors necessarily introduced into the system. To model such errors data driven approaches could potentially be used. However, these approaches have the drawback of having huge numbers of parameters and subsequently being difficult or impossible to estimate for large scale problems.

The continuous time and continuous space approach lends itself particularly well to modeling spatio-temporal problems. This is caused by the continuous time formulation and using infinitesimal time steps. In Paper F we present a framework for spatio-temporal forecasting which is subsequently applied to the problem of producing spatio-temporal forecasts for a solar plant in Nevada, USA. In the following example, Example 3.3.3, we present a spatio-temporal model for solar power forecasting as was developed in Paper F.

Example 3.3.3 Consider the problem of modeling the spatio-temporal dynamics of clouds moving over a solar power facility. This is done with coupled stochastic differential equations in Paper F.

We consider a section of the power plant that consists of a rectangular grid of $I \times J = 5 \times 14 = 70$ inverters. We let the *change* in power output of inverter $[i, j]$ at location $x_{i,j}$ at time t be modeled by the stochastic variable $U_{i,j,t}$. We order the inverters such that inverter $U_{i+1,j,t}$ is the one directly to the east of inverter $U_{i,j,t}$. Also we name the inverters such that $U_{i,j+1,t}$ is the inverter directly north of $U_{i,j,t}$.

The cloud speed measurements are given as input to the model and are denoted by the vector v_t. For modeling purposes, the cloud speed is decomposed into its four directional components, namely North, East, South, and West, denoted by n_t, e_t, s_t, and w_t, respectively. The model proposed is as follows:

$$
\begin{aligned}
dU_{i,j,t} &= \theta|v_t| \Big(\mathrm{n}_t(U_{i,j+1,t} - U_{i,j,t})\mathbb{1}_{\{j+1\leq J\}} + \mathrm{e}_t(U_{i+1,j,t} - U_{i,j,t})\mathbb{1}_{\{i+1\leq I\}} \\
&\quad + \mathrm{s}_t(U_{i,j-1,t} - U_{i,j,t})\mathbb{1}_{\{j-1\geq 1\}} + \mathrm{w}_t(U_{i-1,j,t} - U_{i,j,t})\mathbb{1}_{\{i-1\geq 1\}} \qquad (3.20) \\
&\quad - \mu U_{i,j,t} \left(\mathrm{n}_t \mathbb{1}_{\{j=J\}} + \mathrm{s}_t \mathbb{1}_{\{j=1\}} + \mathrm{e}_t \mathbb{1}_{\{i=I\}} + \mathrm{w}_t \mathbb{1}_{\{i=1\}} \right) \Big) dt \qquad (3.21) \\
&\quad + \sigma dW_{i,j,t} \\
dQ_{i,j,t} &= U_{i,j,t} dt \qquad (3.22) \\
Y_{i,j,k} &= Q_{i,j,t_k} + \epsilon_{i,j,k}, \qquad (3.23)
\end{aligned}
$$

with $\epsilon_{i,j,k} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$. The parameters in the model are thus θ, μ, σ and σ_ϵ, where $\theta|v_t|$ governs the speed at which the value in adjacent cells tend to each other. μ governs how rapidly $U_{i,j,t}$ tends to zero, if $U_{i,j,t}$ is an upwind cell, that is it is on the leading edge towards the wind. σ is the system noise and models perturbations in the system and σ_ϵ characterizes the observation noise. Symbol $\mathbb{1}_{\{\cdot\}}$ represents an indicator or heavyside function, that is equal to 1 if the stated condition is met and 0 otherwise. The indicator functions are used to handle the boundaries of the solar field, such that the model only relates locations that are actually present in the model. This also applies to the dampening term, where we dampen cells on the leading edge towards the wind. Further note that, in this particular case, $Y_{l,k} = Y_{i,j,k}$, where we let l go through all the feasible combinations of $[i, j]$ (that is, model (3.20)–(3.23) assumes that we have power measurements for all locations or inverters, however, this is not a necessary requirement).

If we go to the limit in the grid sizing in the model defined by equations (3.20)-(3.23) a particular simple formulation occurs. We arrive at the stochastic partial differential equation that governs the spatio-temporal dynamics (equation (3.20)):

$$dU(x,t) \quad = \quad \bar{v}\theta\nabla U(x,t)dt + \sigma dW(x,t), \tag{3.24}$$

that is a stochastic unidirectional wave equation describing the cloud movement across the field of solar panels.

As seen from this model formulation, there are only few parameters which makes this model feasible to estimate and to do predictions in a timely manner such that we can provide forecast within the horizons of one time step, that is five seconds.

Figure 3.4 presents the predicted and actual spatio-temporal power output of the before mentioned solar power plant. As seen the drop in power production moves from lower left towards the upper right which is in concordance with the wind direction. We note however, that the cloud speed vector has large errors and these errors accumulate for predictions ranging over longer horizons.

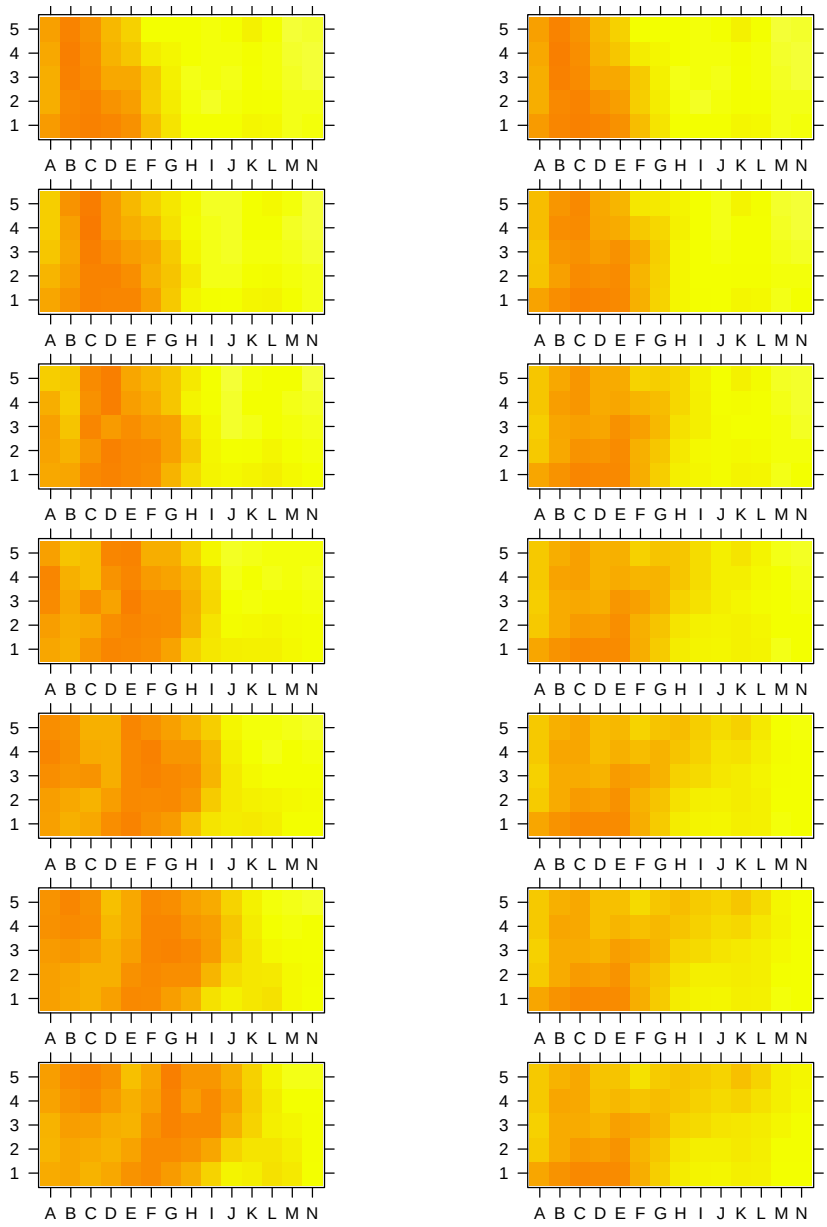

Figure 3.4: *The observed power generation (left) and the predicted power generation (right) from 0 to 60 seconds in 10 second increments.*

Chapter 4

Forecast Products and Applications

It is sometimes argued that forecast mainly serve to comfort decision makers, here the power system participants and operators, while they are not really used or not used in an optimal manner. However, employing appropriate forecasts in a decision making problem can improve the decisions tremendously while similarly allowing for controlling the associated risk. A fundamental point is that all forecasts are to some extent wrong, in the sense that they never predict the future exactly. This should be accounted for in the end implementation and a forecast should not be treated as certain knowledge about the future ([mor]).

No mater the type of forecast, forecasting should always be seen as a form of extrapolation from the present to the future. A model is built and fitted to a set of data, and then used to make predictions on an entirely new data set. Thus forecasts always include, albeit implicitly, *given the information set at our disposal and given the model structure and parameters under the model the prediction about the future is...* As such all forecasts and predictions are conditional on the data used and on the model. Considering this further it does in fact not make sense to calculate expected values of the future without an underlying model. When opinions about the future are expressed, there is always an underlying model, be it qualitative or quantitative.

As discussed in previous chapters forecasting and knowledge about the future have

many applications. In this chapter we will discuss the different forecast products and how they are used. This chapter further highlights the importance of integrating forecasting and decision making for providing the optimal input to decision makers to qualify their decisions. As such forecasting and optimization takes on a new dimension where providing the appropriate forecast products becomes essential for an efficient optimization and where the specific choice of decision making tool must correspond with the information available from the forecast output. We stress these facts by going through examples of how forecasts are used in energy systems applications in different ways. For further discussion on how forecasts are applied optimally to power systems and electricity markets, we refer the interested reader to [mor].

4.1 Point Forecasts

When the renewable energy generation forecast at time t for time $t + k$ is a single value, the forecast is referred to as a point prediction or point forecast. Point forecasts are what most people think of when they think of a forecast. The point forecast from a physical model will look similar to that of a data driven model or gray box model. While in physical or deterministic models the forecast uncertainty is to some extent ignored, for data-driven models this is not the case, and as such the point forecast is typically the expected value or most likely outcome.

Applications of point forecasts range from market operation of renewable generation, in grids where there is little renewable penetration to electricity or heat load forecasting. In Example 4.1 we consider the point forecasts that are outputs of the gray box model for wind power generation for the Klim wind farm developed in Paper E.

Example 4.1 An example of a point forecast is the multi-horizon forecast for the Klim wind farm discussed in previous examples. The point forecast is generated from the model presented in Paper E. The forecast for a horizon of 1-48 hours is presented in Figure 4.1 where also the realizations of produced power are shown.

4.2 Probabilistic Forecasts

As point predictions provide a single value for each time point within the prediction horizon, they neglect what *could* happen which is often crucial for operational problems, where the costs incurred are affected by the whole range of potential outcomes.

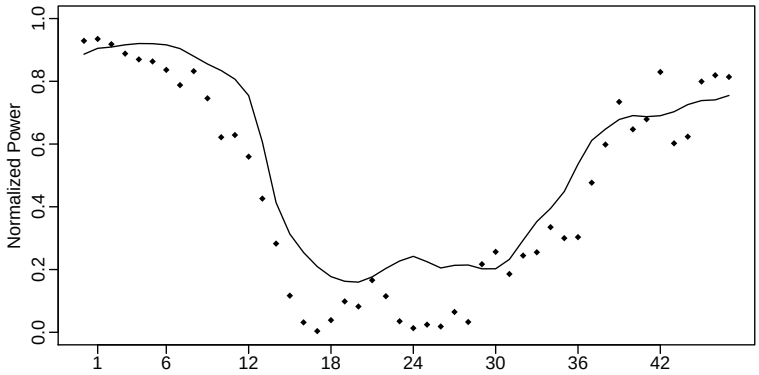

Figure 4.1: *The point predictions of generated power from the Klim wind farm for horizons 1-48 hours ahead, the black line, and the actual realizations, the black dots.*

Hence, development of models that can produce a range of outputs have received substantial attention.

A probabilistic forecast issued at time t for time $t+k$ consists of a predictive probability density function (pdf) or some summary statistic from this pdf for the underlying stochastic variable. We can therefore interpret point forecasts as probabilistic forecasts with point mass of 1 in the predicted value. We go into some of the most commonly used summary statistics for probabilistic forecasting.

4.2.1 Quantile Forecasts

Quantile forecasts are based on the predictive density forecast. A quantile forecast of level α at time t, denote this by $q^{(\alpha)}_{t+k|t}$, has the information that at some future time $t+k$ that there is a probability of α that the forecast variable will be less than $q^{(\alpha)}_{t+k|t}$.

In Figure 4.2 the $q^{(0.25)}_{t+k|t}$ quantile forecast for the normalized power generation is shown. This is obtained from the model developed in Paper E.

Applications of quantile forecasts range from market trading strategies (as seen in Example 2.1) for renewable energy generation to quantile forecasts of wind speeds for maintenance scheduling, to selecting the proper reserve size.

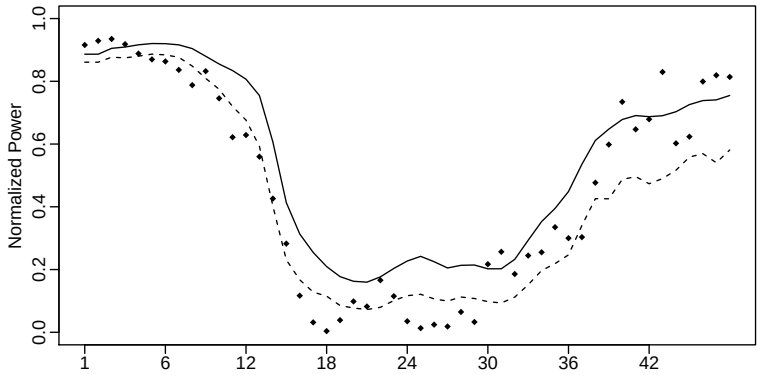

Figure 4.2: *The 25 % lower quantile of generated power from the Klim wind farm for horizons 1-48 hours ahead, the dashed line, the point prediction, the black line, and the actual realizations, the black dots.*

4.2.2 Prediction Intervals

Quantile forecasts describe probabilistic aspects about the future on a threshold basis. Even though this is directly relevant for a large set of operational problems, receiving a single quantile forecast can leave end users with doubt about the level of forecast uncertainty for the coming period. Prediction intervals alleviate this issue by supplying in interval that the variable in question is going to be with an with some level of certainty.

The prediction interval of level β issued at time t for time $t+k$, $I^{(\beta)}_{t+k|t}$, defines a range of possible values such that the probability of the forecasted variable is within this interval with probability β. Immediately it should become clear that there is a link with quantiles. Prediction intervals defined in this way are not unique as some of the probability mass may be shifted form one tail to the other. For this reason we define the symmetric or central prediction interval with and equal mass in each tail. Thus we get that:

$$I^{(\beta)}_{t+k|t} = \left[q^{(\frac{\beta}{2})}_{t+k|t}, q^{(1-\frac{\beta}{2})}_{t+k|t} \right] \tag{4.1}$$

In Figure 4.3 we see the 75 % symmetric prediction intervals for the Klim wind farm.

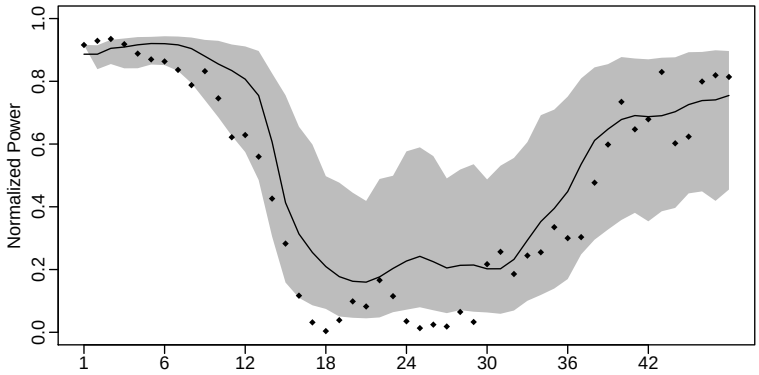

Figure 4.3: *The 75 % symetric prediction interval of generated power from the Klim wind farm for horizons 1-48 hours ahead, the gray shaded area, the point prediction, the black line, and the actual realizations, the black dots.*

4.2.3 Predictive Densities

Point predictions, quantile forecasts and prediction intervals are in fact only part of describing the whole information about the future at time $t + k$. This whole information at time $t + k$ is contained in the predictive density for time $t + k$ in the future given by $f_{t+k|t}$. Here $f_{t+k|t}$ is the density function describing the distribution of the variable in question.

While predictive densities contain the most information of the forecasts discussed so far, they require the forecast user to have knowledge of statistics and probability theory in order to be used properly. The predictive density may well be skewed and time varying and thus require a knowledgeable interpreter.

Predictive densities are used in many decision making problems related to renewable energy management and their integration into electricity markets. Among these applications are optimal bidding strategies as well as grid reserve requirements that account for the entire uncertainty in the electrical supply. Details of their use can be found in [mor].

4.3 Multi-Horizon Forecasts

Probabilistic forecasts provide substantial insight into the characteristics of the stochastic process in question. However, they allow us only to consider the marginal densities for each lead time. What is not considered is the temporal interdependence between

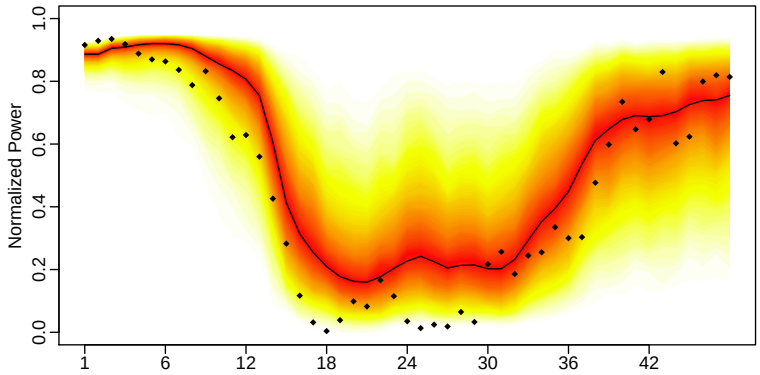

Figure 4.4: *The predictive distribution of generated power from the Klim wind farm for horizons 1-48 hours ahead, warmer colors indicate a higher probability, the point prediction, the black line, and the actual realizations, the black dots.*

different lead times. Considering Figure 4.4 it becomes clear that there is a large auto-correlation in the forecast errors. This means that if a large forecast error is observed at time $t + k$ it is likely followed by a large forecast error with the same sign at time $t + k + 1$. Such information about the temporal dependence structure in time (and space for forecasts distributed spatially) may be crucial for a number of operational problems resulting from an integrated management and specific physical constraints. These include the operation of a virtual power plant consisting of a energy storage devise and a wind farm, unit commitment problems and in systems with a large renewable component where physical restrictions may apply to the ramp rates of power generators.

4.3.1 Scenarios

Presenting a fan of scenarios serves to capture some of the inter-temporal dependence in the forecasts. These forecast can be generated by simulating the underlying model.

In Figure 4.5 we present simulated trajectories for the wind speed observed at the Klim wind farm. Notice here that the simulated trajectories show the same type of auto correlation in the forecast errors as observed in the data. Thus the scenarios capture some of this inter-temporal dependence in the forecasts.

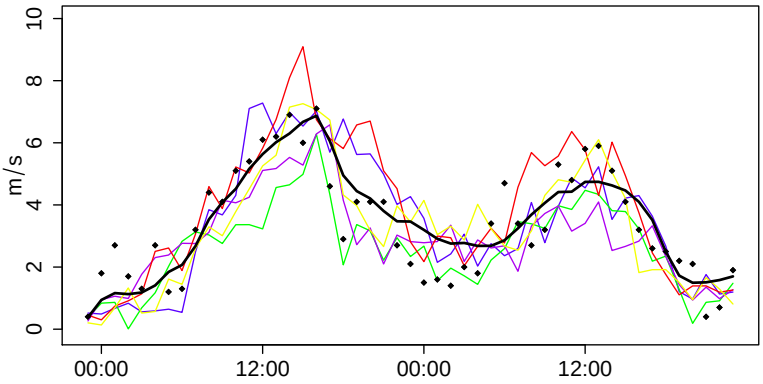

Figure 4.5: *Simulated scenarios for 1-48 hours ahead of wind speed, the colored lines, the prediction from the model, the black line, and the realized observations, black dots.*

4.3.2 Conditional Predictive Densities

To further increase the information available conditional predictive densities can be provided. We denote this conditional predictive density by $\tilde{f}_{t+k|t+j}$, which is the prediction of the future at time $t+k$ given the yet unknown state of the system at time $t+j$, where $j < k$. Thus for each possible value of the the system at time $t+j$ we get a specific predictive density function for time $t+k$.

Conditional predictive densities are used in many operational aspects of energy systems e.g. electricity trading with several gate closures, say day-ahead market and the regulation market. In this case first the predictive density of generated power for the time $t+k$ is used to construct a bid into the market at the time of the first gate closure, time t. Subsequently at the time of the gate closure of the regulation market, at time $t+j$ with $j < k$, a bid for up or down regulation is placed. In order to bid in an optimal manner the predictive density, $f_{t+k|t}$, along with the conditional predictive density, $\tilde{f}_{t+k|t+j}$, are needed.

4.4 Output Model

The final forecast product that can be used for decision making problems is when the entire predictive model is seamlessly integrated into the decision making framework. We denote such a forecast product M_t, where there is a one-to-one relation between the forecast and the underlying model. While in general this is not feasible due to

numerical considerations, there are specific cases, where such integration is possible. An example of this is the model for driving patterns developed in Paper A that is input into the optimal charging scheme for an electric vehicle presented in B, which we show in brief in Example 4.4.

Example 4.4 We build upon the model for driving patterns of a single vehicle presented in Paper A. The model for driving patterns is based on a time inhomogeneous Markov chain which is completely characterized by the time changing transition probabilities. Thus there is a one-to-one relationship between the transition probabilities and the underlying model. In Paper B these transition probabilities are input into the users objective that is to minimize the costs of charging the electric vehicle while maintaining enough charge on the vehicle such that it is still able to service the driving needs. The users objective function is defined as the expected revenue (negative costs) from time t to T:

$$J_t^\pi(S_t) = \mathbb{E}\left[\sum_{\tau=t}^{T} R_\tau(S_\tau, u_\tau^\pi(S_\tau)) \middle| S_t\right]. \tag{4.2}$$

Let $\mathcal{U}_s$ denote the set of feasible decisions. A policy, π, is a collection of decisions $u_t^\pi(s) \in \mathcal{U}_s$, spanning the horizon from $t = 0$ to $t = T$ and all states S_t. Let Π denote the set of all feasible policies. Hence, for each t and each state S_t, $\pi \in \Pi$ will contain the action, $u_t^\pi(S_t)$, under the policy π. $R_\tau(S_\tau, u_\tau^\pi(S_\tau))$ is the revenue at time τ, in state S_τ under the policy π_τ. We note here that the objective function also takes a predefined input that is the penalty function.

From looking at the objective function in equation (4.2), the probabilities enter into the objective function in how the expected value is computed. The model makes use of the direct probabilities and thus the optimization framework makes use of the predictive model in its entirety

The objective is then to find a policy, π^*, that satisfies:

$$J_t^{\pi^*}(S_t) = \sup_{\pi \in \Pi} J_t^\pi(S_t), \tag{4.3}$$

for all $0 \leq t \leq T$. The optimal solution is subsequently found by stochastic dynamic programming using Bellmans principle of optimality.

Solving the problem of finding the optimal charging policy for specific penalty levels yields different charging strategies. Some realized charging strategies are shown in Figure 4.6. From this figure it is seen that having a higher penalty for not servicing the driving needs results in keeping the average state of charge on the battery higher. Hence the charging of the vehicle is not postponed to the same extent as for lower penalties.

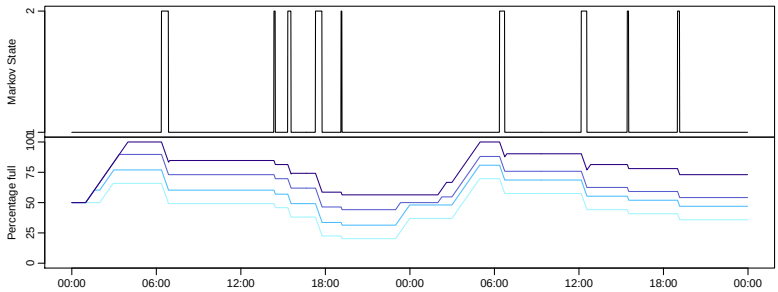

Figure 4.6: *Top: realization of the driving pattern. Bottom: the corresponding charge on the battery in percent for different penalty values when implementing $\tilde{\pi}^*$. The lightest blue line refers to the lowest penalty and the penalty increases with the darker shades of blue.*

Chapter 5

Conclusion and Perspectives

5.1 Conclusion

Renewable energy generation has seen a rapid growth in recent years. This challenges the existing power system infrastructure by introducing a large quantity of uncertain power generation into the generation mix. In order to integrate this renewable energy in an efficient manner a variety of initiatives have been taken and more need to follow as the renewable energy penetration increases. What remains a common factor for an efficient management of this uncertain power generation is that knowledge about the immediate future greatly improves decision makers ability to make good decisions. This knowledge of the future is materialized as forecasts of renewable energy generation, energy demand response, power loads and power prices. High quality forecasts for the energy system are required. For efficient integration it is important that forecasters and decision makers acknowledge that they have an integrated challenge and both perspectives must be considered in unison.

This dissertation focus on developing models for forecasting applied to energy systems and to relate them to optimal decision making. We have focused on models that are a mix between purely physical and purely data driven models known as gray box models. This has been used to impose physical real-world constraints into the models that allow for better predictions, an interpretable physical relationship and a large reduction in the number of parameters. In doing so we have developed new ap-

proaches to energy systems forecasting where we use a continuous time perspective to substantially simplify models and to ensure that the models behave as expected. Improvements to the state-of-the-art have been observed in various applications ranging from describing vehicle driving patterns to solar irradiance forecasts to wind power forecasts. The improvements have been observed in diverse forecasts products including: point forecasts, quantile forecasts, predictive densities and scenario generation.

Starting out from the perspective of the end user a model for the driving patterns of a single vehicle is developed in Paper A and further developed into a model for optimal charging of an electric vehicle weighing the cost of charging against vehicle availability in Paper B. These papers take the point of view of the single end user, believing that energy systems operation, at least on the small scale, should be managed by indirect control, where a price signal or similar is sent to end users. This allows the end users to control their energy consumption at their leisure, albeit aided by smart algorithms and automation that they control. This stands in contrast to direct control, where a central entity manages the end-users energy consumption. In the case of electric vehicle charging this would typically result in scheduling the charging of electric vehicles not considering the individual demand. Paper G consider the economic impact of various vehicle charging regimes when there is a varying and unknown power price.

A model framework for continuous sample space/continuous time forecasting is presented in Paper C, where it is applied to solar irradiance forecasting. Here the state-of-the-art is advanced by a method to provide reliable probabilistic forecast of solar irradiance accurately describing the predictive densities which are confined to the actual possible outcomes with non-Gaussian distributions of the predicted variable. This formulation is made possible by the continuous time nature of the model that reflects the true dynamics more accurately than typical discrete time models. Hence the model out-performs several benchmarks. The methodology presented in Paper C is further developed in Paper D which provides a model that can produce accurate probabilistic forecasts of wind speeds and out-performs benchmark models. Integrating numerical weather predictions into gray box models are vital for accurate forecasts on longer time horizons. The methodology for doing this is further developed in Paper D. Paper E, using the same basic methodology as in Papers C and D, advances the state-of-the-art for wind power forecasting by building on the probabilistic wind speed model in Paper D and introducing an adaptive power curve for translating wind speeds into wind power. Taking advantage of an understanding of the underlying physics of the system allows for producing better forecasts and scenarios and accurately capturing the auto-correlation in forecast errors for all time horizons and outperforms state-of-the-art benchmark models. Thus a physical understanding of the system at hand may greatly help improve forecast performance.

The problem of producing online spatio-temporal forecast, that is forecasts for multiple spatial locations as well as for multiple lead times, is highly relevant to energy

systems application but presents immense challenges. Classical approaches use physical models, but thereby neglect correlation structures, predictive densities and space-time scenarios. Data driven models have the potential of solving these shortcomings but this comes at the expense of the models becoming very expensive computationally with infeasible number of parameters. In Paper F we presented a probabilistic spatio-temporal gray box model based on an understanding of the underlying physics. This allowed us to have an online model with few (four) parameters, at 70 locations at sampling rate of 5 seconds and allow for predictions multi-steps ahead within this sampling frequency. The model outperforms state-of-the-art benchmarks and allows for producing predictive densities, correlations and space-time scenarios.

5.2 Perspectives

This thesis has given rise to a number of different questions and ideas for future work. In this section we do not aim to discuss the structure of the power system in the far future but rather a pragmatic approach discussing the next steps for improving and extending the methodology and applications.

Demand response is seen as a crucial piece for integrating large amounts of renewable energy generation. To do this in a manner acceptable for end users their consumption needs to be accounted for as done in Paper A and B for electric vehicles. The methodology presented in these two papers are, however, not limited to the charging of electric vehicles. Obvious extensions are power or energy consumption where the appliance has discrete usage states (or close approximations hereof). Examples could be the smart water heater, with the users having different water consumption states, say showering, dish-washing and so on. Other applications could be occupancy of a house, which is either occupied or not and can adjust the heating accordingly. An obvious extension to the driving patterns model is to consider the model as part of a more general population. This allows for studying the behavior of multiple vehicles and their effect on the power grid. In a more general setting this could also be used to model demand response, from a bottom up approach.

The methodology presented in Papers C, D and E can potentially be used for forecasting other aspects of power systems operations, such as power load, heat load and power prices. A further extension is the nature of the continuous time framework that allows for multidimensional forecasts, that is predicting e.g. both power production and prices in the same model. A similar thing is already done in Paper E, where wind speeds and wind power were forecast by the same model. This allows for a better description of the uncertainty and may effect power systems operation. A final model could include different types of generation within an area, the different loads and possible interconnections to other areas. As continuous time models can

seamlessly switch between different sampling frequencies it could also possibly be a fertile endeavor to go into the models on different time scales. Potentially this would also help produce better models as the model parameters should be invariant of the sampling frequency. Fitting models to different time scales could potentially improve the generated forecasts. Scoring the different forecasts becomes important as to also acknowledge more information than the state-of-the-art scores that focus on predictive densities. What is also needed are scores that can distinguish how well scenarios are generated by the model and how well the time-interdependence is captured.

A wide range of spatio-temporal problems are of interest to the energy community. Future work could be directed at using the methodology developed in Paper F to solve problems such as forecasting distributed generation from wind farms and distributed solar power generation. This would extend the work considered in Paper F to a possible no-regular grid an develop the framework on longer time scales. Future work could also be directed at modeling the demand side of the energy system to accurately forecast spatio-temporal demand, possibly with the addition of distributed solar power, which will introduce a weather dependence in the demand. Extensions of the work done in Paper F could also focus on modeling the noise process as opposed to primarily considering the deterministic part. As different solutions to space-time problems are posed it also becomes important to develop scoring rules for space-time problems, which are as of yet underdeveloped. However, it should be stressed that the derivation of new scores is not a trivial task as one needs to ensure that the proposed scoring rules are proper.

Bibliography

[Age14] International Energy Agency. *World Energy Outlook 2014*, volume 1. International Energy Agency, 2014.

[Ati01] Amir F Atiya. Bankruptcy prediction for credit risk using neural networks: A survey and new results. *Neural Networks, IEEE Transactions on*, 12(4):929–935, 2001.

[BHBG95] Radmila Bubnová, Gwenaëlle Hello, Pierre Bénard, and Jean-François Geleyn. Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/Aladin NWP system. *Monthly Weather Review*, 123(2):515–535, 1995.

[Bjö09] T. Björk. *Arbitrage Theory in Continuous Time*. Oxford Finance Series. OUP Oxford, 2009.

[CCVO98] W Charytoniuk, M-S Chen, and P Van Olinda. Nonparametric regression based short-term load forecasting. *Power Systems, IEEE Transactions on*, 13(3):725–730, 1998.

[CGS08] Michael J Cooper, Huseyin Gulen, and Michael J Schill. Asset growth and the cross-section of stock returns. *The Journal of Finance*, 63(4):1609–1651, 2008.

[DATD04] Ioannis G Damousis, Minas C Alexiadis, John B Theocharis, and Petros S Dokopoulos. A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *Energy Conversion, IEEE Transactions on*, 19(2):352–361, 2004.

[fra] Electricity production from solar and wind in germany in 2014, author=Fraunhofer ISE, volume=1, year=2014, publisher=Fraunhofer ISE.

[GDS+94] Georg A Grell, Jimy Dudhia, David R Stauffer, et al. A description of the fifth-generation Penn State/NCAR mesoscale model (mm5). Technical report, Penn State, 1994.

[HBG04] Lars Hufnagel, Dirk Brockmann, and Theo Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America*, 101(42):15124–15129, 2004.

[Hov14] Kjetil Malkenes Hovland. Denmark's wind power output rises to record in first half. *The Wall Street Journal*, 2014.

[K96] E. Källén. Hirlam documentation manual; 1996. system 2.5. Technical report, Norrkøping, Sweden, 1996.

[Kri12] Tarjei Kristiansen. Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy*, 49:328–332, 2012.

[KZS09] Andrew Kusiak, Haiyang Zheng, and Zhe Song. Short-term prediction of wind farm power: a data mining approach. *Energy Conversion, IEEE Transactions on*, 24(1):125–136, 2009.

[LD03] Jonathan M Levine and Carla M D'Antonio. Forecasting biological invasions with increasing international trade. *Conservation Biology*, 17(1):322–326, 2003.

[LGS+08] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, and G. Kallos. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(12):2348 – 2362, 2008.

[LSS14] Yanting Li, Yan Su, and Lianjie Shu. An armax model for forecasting the power output of a grid connected photovoltaic system. *Renewable Energy*, 66:78–89, 2014.

[mor] *Integrating Renewables in Electricity Markets, author= Morales, Juan M. and Conejo, Antonio J. and Madsen, Henrik and Pinson, Pierre and Zugno, Marco.*

[NNM+07] Henrik Aa Nielsen, Torben S Nielsen, Henrik Madsen, Maria J Pindado, and Ignacio Marti. Optimal combination of wind power forecasts. *Wind Energy*, 10(5):471–482, 2007.

[Øks10] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext (1979). Springer, 2010.

[PCD+01] Kenneth Patterson, Robin Cook, Chris Darby, Stratis Gavaris, Laurence Kell, Peter Lewy, Benoît Mesnil, André Punt, Victor Restrepo, Dankert W Skagen, et al. Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries*, 2(2):125–157, 2001.

[PM12] Pierre Pinson and Henrik Madsen. Adaptive modelling and forecasting of offshore wind power fluctuations with markov-switching autoregressive models. *Journal of Forecasting*, 31(4):281–313, 2012.

[PN06] Cameron W Potter and Michael Negnevitsky. Very short-term wind forecasting for tasmanian power generation. *Power Systems, IEEE Transactions on*, 21(2):965–972, 2006.

[PPP13] Marta Poncela, Pilar Poncela, and José Ramón Perán. Automatic tuning of kalman filters by maximum likelihood methods for wind energy forecasting. *Applied Energy*, 108:349–362, 2013.

[Rei09] Gordon Reikard. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83(3):342–349, 2009.

[Sad06] Perry Sadorsky. Modeling and forecasting petroleum futures volatility. *Energy Economics*, 28(4):467–488, 2006.

[SKD+05] William C Skamarock, Joseph B Klemp, Jimy Dudhia, David O Gill, Dale M Barker, Wei Wang, and Jordan G Powers. A description of the advanced research wrf version 2. Technical report, DTIC Document, 2005.

Part II

Papers

Paper A

Inhomogeneous Markov Models for Describing Driving Patterns

Submitted to *IEEE Transactions on Smart Grid, 2015.*

Inhomogeneous Markov Models for Describing Driving Patterns

Emil B. Iversen, Jan K. Møller, Juan M. Morales, *Member, IEEE,*, and Henrik Madsen

***Abstract*—It has been predicted that electric vehicles will play a crucial role in incorporating a large renewable component in the energy sector. If electric vehicles are integrated in a naive way, they may exacerbate issues related to peak demand and transmission capacity limits while not reducing polluting emissions. Optimizing the charging of electric vehicles is paramount for their successful integration. This paper presents a model to describe the driving patterns of electric vehicles, in order to provide primary input information to any mathematical programming model for optimal charging. Specifically, an *inhomogeneous Markov model* that captures the diurnal variation in the use of a vehicle is presented. The model is defined by the time-varying probabilities of starting and ending a trip and is justified due to the uncertainty associated with the use of the vehicle. The model is fitted to data collected from the actual utilization of a vehicle. Inhomogeneous Markov models imply a large number of parameters. The number of parameters in the proposed model is reduced using B-splines.**

***Index Terms*—B-splines, , Driving patterns, Electric vehicles, Inhomogeneous Markov chain, Hidden Markov model**

I. Introduction

Electric vehicles (EVs) have no emissions and are a sustainable alternative to conventional vehicles, provided that the energy used for charging is generated by renewable sources. Electricity generation from renewable energy sources, such as wind, solar and wave energy depends on weather conditions and consequently, is inherently stochastic. Still in the absence of a large-scale infrastructure for energy storage, today electricity has to be produced and consumed at the same time. With electricity coming from renewable sources it is not possible to produce additional power, if weather conditions do not allow for it. Moreover, in times of high availability from renewable sources, the demand for power may be low and the economic potential of renewables may thus be wasted. Electric vehicles may help overcome this issue by charging the EVs when energy from renewable sources is abundant and by supplying power into the electrical grid at times of high demand. As long as electric vehicles are charged with electricity from renewable sources, they represent a sustainable zero-emissions alternative to conventional fossil-fuel-based vehicles. On the contrary, if EVs are charged in a naive way, they may increase the peak electricity demand. As a consequence, the extra energy needs would have to be covered by peak-supply units, typically fossil-fuel-based, which would nullify the decrease in emissions gained by switching from conventional to electric vehicles. Furthermore, an increased peak demand could lead to a shortage of transmission capacity, which would force an expansion of the electrical grid to handle the higher peak demand. This is costly and undesirable. To avoid these problems, EVs should be charged in a smart fashion.

As EVs are primarily used for transportation, and not for energy storage, it is essential to charge each vehicle such that there is enough energy to cover any desired trip. Hence a decision-support tool is required to determine whether it is possible to postpone the EV charging or whether it should be charged right away. For such a tool to produce optimal charging decisions, a model capturing the utilization of a specific vehicle is essential. The complexity of human behavior calls for a stochastic model to adequately describe the driving needs of EV users.

In the technical literature, albeit observed vehicle usage has been considered in several studies ([1], [2]), the stochastic modeling of the driving patterns of a single vehicle has received little attention ([3]). Rather, the scientific community has been more focused on both the analysis of the potential impact of charging EVs and the design of models to decide when to charge ([4]). In this vein, the effect of the large-scale integration of EVs into the power grid has been studied in several papers, ([5], [6], [7], [8], [9]). Issues such as peak load, different charging strategies, network losses, minimizing costs and market equilibrium strategies have been considered.

All that said, the idea of using inhomogeneous Markov chains to model EVs usage is not new. One the one hand, a number of authors employ this modeling approach to simulate the utilization of a population of EVs with the aim of characterizing the total electricity demand of the whole fleet ([10], [11], [12], [13], [14]). However, due to practical and methodological issues, these works prove to be too coarse when it comes to capturing the use of a *single* vehicle. Common for these works is that they base the modeling of the aggregate behaviour of the EV fleet on the assumption that the same underlying stochastic process generates the driving patterns of all the vehicles in the population. Even though this approximation may be accurate enough for applications where the description of the aggregate behavior of the EV fleet is all that is needed (e.g. to estimate the electricity consumption of the fleet), it is not good enough for applications where the modeling of the driving patterns of a specific vehicle is required (e.g., for smart charging of an EV, where knowing, for instance, that people on average drive to work at 7:30 am is not that useful to the individual EV owner). This is so because the stochastic dynamics of the utilization of a particular vehicle can be very different from the stochastic dynamics of the population. Therefore, modeling efforts should concentrate on data pertaining to that specific vehicle. Furthermore, from a methodological point of view, modeling the driving behavior of a single vehicle poses nontrivial challenges related to

limited data, validity of the Markov assumption on the trip duration, and requirements on time resolution, all of which are properly addressed in this paper. On the other hand, Markov models have also been used to simulate the utilization of a single vehicle while driving, focusing on trip duration, consumption, speed and type-of-use, ex. highway, rural or urban ([15], [16]). These works, however, focus solely on the driving and thus disregard the time-varying use of the vehicle, which is essential in practical applications, such as smart charging.

This paper bridges the gap between these two approaches and presents a genuine single vehicle model. The model can be easily exploited, for example, by decision-making tools for charging an EV. Furthermore, our model provides advances in modeling driving patterns and does not rely on the typical, average or stylized use of a vehicle. The model is fitted to a specific vehicle based on observed data from the utilization of that vehicle. An *inhomogeneous Markov model* is applied to capture the diurnal variation of the driving pattern. A major disadvantage of these types of models is the high number of parameters to be estimated. A generalized linear model that makes use of B-splines is then applied to substantially reduce this number, while overcoming the problem of missing observed transition states due to limited data and the need for a high temporal resolution. An algorithm is proposed to place knots and to find the appropriate number of knots needed for the B-splines. We use hidden Markov models to adequately capture the duration of the trips. The proposed model does not rely on any assumptions regarding the use of the vehicle, and consequently a versatile model is obtained. Applying the model within a stochastic optimization framework will allow for capturing issues related to charging, availability, and costs of using an EV.

The paper is organized as follows: Section 2 gives a brief introduction to inhomogeneous Markov chains. In Section 3 the number of parameters in the model is reduced by applying B-splines to a generalized linear model. Section 4 provides a numerical example of the model, where the parameters are fitted to observed data from a single vehicle. Section 5 concludes and provides directions for future research within this topic.

II. An Inhomogeneous Markov Chain

A state-space approach is proposed to describe the use of a vehicle. This approach models the vehicle as being in one of several distinct states. In its simplest form the model has two states, which capture whether the vehicle is either *driving* or *not driving*. A more extensive model may include information about where the vehicle is parked, where it is driving, or what type of trip the vehicle is on. In this section we start from a general state-space approach and finish with a detailed description of the two-state model.

A. Discrete Time

Consider a sequence X of random variables X_t, $t \in \{0, 1, 2, \ldots\}$, which take on values in the countable set S, referred to as the state space. Without loss of generality, we assume that the state space includes N states. A Markov chain is a random process where future states, conditioned on the present state, do not depend on the past states ([17]). In discrete time, $\{X\}$, is a Markov chain if

$$\mathbb{P}\left(X_{t+1} = k | X_0 = x_0, \ldots, X_t = x_t\right) = \mathbb{P}\left(X_{t+1} = k | X_t = x_t\right) \tag{1}$$

for all $t \geq 0$ and all $\{k, x_0, \ldots, x_t\} \in S$.

A Markov chain is uniquely characterized by the transition probabilities, i.e.

$$p_{jk}(t) = \mathbb{P}\left(X_{t+1} = k | X_t = j\right). \tag{2}$$

If the transition probabilities do not depend on t, the process is called a homogeneous Markov chain. If the transition probabilities depend on t, the process is known as an inhomogeneous Markov chain.

Considering the use of a vehicle, it is reasonable to expect that the probability of a transition from state j to state k at any specific weekday is the same. Thus the transition probabilities on Tuesday in one week are assumed to be the same as on Tuesdays in other weeks. Furthermore, it is natural to assume that the transition probabilities are the same on all weekdays, that is, from Mondays to Fridays. If the sampling time is in minutes, this leads to the assumption:

$$p_{jk}(t) = p_{jk}(t + 1440), \tag{3}$$

where 1440 is the number of minutes in a day. In other words the transition probabilities, defined by (2), are constrained to be a function of the time s in the diurnal cycle. The matrix containing the transition probabilities is given by

$$\mathbf{P}(s) = \begin{pmatrix} p_{11}(s) & p_{12}(s) & \ldots & p_{1N}(s) \\ p_{21}(s) & p_{22}(s) & \ldots & p_{2N}(s) \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1}(s) & p_{N2}(s) & \ldots & p_{NN}(s) \end{pmatrix}, \tag{4}$$

where $p_{jj}(s) = 1 - \sum_{i=1, i \neq j}^{N} p_{ji}$.

If the model is formulated with time resolution in minutes, $s \in \{1, 2, \ldots, 1440\}$. The assumed periodicity from (3) implies that all the observations from different days are lumped together and a transition is only denoted by its time of day. It follows that the conditional likelihood function, for the model with N states, is given by ([18]):

$$L\left(\mathbf{P}(1), \mathbf{P}(2), \ldots, \mathbf{P}(1440)\right) = \prod_{s=1}^{1440} \prod_{j=1}^{N} \prod_{k=1}^{N} p_{jk}(s)^{n_{jk}(s)}, \tag{5}$$

where $n_{jk}(s)$ is the number of observed transitions from state j at time s to state k at time $s + 1$, where s is the time in minutes of the diurnal cycle.

From the conditional likelihood function the maximum-likelihood estimate of $p_{jk}(s)$ can be found as:

$$\widehat{p}_{jk}(s) = \frac{n_{jk}(s)}{\sum_{k=1}^{N} n_{jk}(s)}. \tag{6}$$

A discrete time Markov model can be formulated based on the estimates of $\mathbf{P}(1), \mathbf{P}(2), \ldots, \mathbf{P}(1440)$. One apparent

disadvantage of such a discrete time model is the huge number of parameters, namely $N \cdot (N-1) \cdot 1440$, where $N \cdot (N-1)$ parameters have to be estimated for each time step. Needless to say, the number of parameters to be estimated increases as the number of states increases. Another problem is linked to the number of observations, i.e. if $\sum_{k=1}^{N} n_{jk}(s') = 0$ for some s', then $\widehat{p}_{jk}(s)$ is undefined.

A reduction in parameters may be obtained if the diurnal variation is negligible for some transitions, i.e. $p_{jk}(s)$ does not depend on s for some pair $\{j,k\}$.

One way to reduce the parameters is to increase the time between samples. If the sampling time is every 10 minutes, the number of parameters would decrease to $N \cdot (N-1) \cdot 144$. This approach is a bit coarse and the number of parameters is still large. Besides, if another parameter reduction technique is subsequently applied to the data, information is lost compared to directly applying the technique to the data with a sampling time in minutes.

In the model with only two states, namely *driving* and *not driving*, the one-minute transition probability matrix becomes:

$$\begin{aligned} \mathbf{P}(s) &= \begin{pmatrix} p_{11}(s) & p_{12}(s) \\ p_{21}(s) & p_{22}(s) \end{pmatrix} \\ &= \begin{pmatrix} 1-p_{12}(s) & p_{12}(s) \\ p_{21}(s) & 1-p_{21}(s) \end{pmatrix}. \end{aligned} \tag{7}$$

The number of parameters is then $2 \cdot 1440$. Assuming that the duration of the trip does not depend on the time of the day, i.e. $p_{21}(s) = p_{21}$, (with 2 being "*driving*" and 1 "*not driving*") the number of parameters is reduced to $1440 + 1$. Note that, as a result of this reduction, the duration of a trip is captured by a single parameter.

It follows that the conditional likelihood function, for the model with two states, is given by:

$$L\left(\mathbf{P}(1), \mathbf{P}(2), \ldots, \mathbf{P}(1440)\right) = \prod_{s=1}^{1440} \prod_{j=1}^{2} \prod_{k=1}^{2} p_{jk}(s)^{n_{jk}(s)}, \tag{8}$$

and the maximum-likelihood estimate $\hat{p}_{jk}(s)$ is computed from (6).

B. Continuous Time

The continuous time analog to the discrete time inhomogeneous Markov chain is presented below. The continuous time version provides a parameter reduction over the discrete time version, if certain structures are present and can be identified. Specifically, if the number of states is larger than two and it is impossible to switch directly between certain pairs of states, the continuous time variant will lead to a parameter reduction. Hence, if such structures are present, the continuous time variant is preferred over the discrete time model. To introduce the continuous time inhomogeneous Markov chain, we define ([17]):

$$p_{jk}(t,u) = \mathbb{P}(X(u) = k | X(t) = j), \tag{9}$$

where $t < u$. The model is based on the following assumptions when $\Delta u \to 0$:

$$p_{jj}(u, u+\Delta u) = 1 - q_{jj}(u)\Delta u + o(\Delta u) \tag{10}$$

$$p_{jk}(u, u+\Delta u) = q_{jk}(u)\Delta u + o(\Delta u) \quad \forall j \neq k, \tag{11}$$

also $0 \leq q_{jj}(u) < \infty$ and $0 \leq q_{jk}(u) < \infty$. The $q_{jk}(u)$'s are known as the transition intensities. These assumptions lead to Kolmogorov's forward differential equation for inhomogeneous Markov processes, expressed in matrix notation as:

$$\frac{\partial \mathbf{P}(t,u)}{\partial u} = \mathbf{P}(t,u)\mathbf{Q}(u) \tag{12}$$

where $\mathbf{P}(t,u) = \{p_{jk}(t,u)\}$, i.e. $\mathbf{P}(t,u)$ is the matrix containing the $p_{jk}(t,u)$'s. The matrix of transition intensities then becomes:

$$\mathbf{Q}(u) = \begin{pmatrix} -q_{11}(u) & q_{12}(u) & \cdots & q_{1N}(u) \\ q_{21}(u) & -q_{22}(u) & \cdots & q_{2N}(u) \\ \vdots & \vdots & \ddots & \vdots \\ q_{N1}(u) & q_{N2}(u) & \cdots & -q_{NN}(u) \end{pmatrix}. \tag{13}$$

Since $\sum_{k=1}^{N} p_{jk}(u, u+\Delta u) = 1$, it follows from (10)-(11) that $\sum_{k=1}^{N} q_{jk}(u) = 0 \ \forall \ j$, i.e. $q_{jj}(u) = \sum_{k=1, k\neq j}^{N} -q_{jk}(u) \ \forall \ j$.

A simple Kolmogorov's differential equation is obtained if $\mathbf{Q}(t)$ is constant in the period $[t, t+T]$:

$$\mathbf{P}(t, t+T) = e^{\mathbf{Q}(t)T}\mathbf{P}(t,t) = e^{\mathbf{Q}(t)T}, \tag{14}$$

where $\mathbf{P}(t,t)$ contains the probability of moving between the different states between t and t, i.e. in zero time, which is a matrix with ones on the diagonal and zero everywhere else. Suppose that $T = 1$. Then the one minute transition probabilities are given by:

$$\mathbf{P}(t, t+1) = \mathbf{P}(t) = e^{\mathbf{Q}(t)}, \tag{15}$$

where $\mathbf{P}(t)$ is the standard transition probability matrix for a discrete time Markov chain. If the model has two states, the matrix of transition intensities becomes:

$$\mathbf{Q}(u) = \begin{pmatrix} -q_{11}(u) & q_{12}(u) \\ q_{21}(u) & -q_{22}(u) \end{pmatrix} = \begin{pmatrix} -q_{12}(u) & q_{12}(u) \\ q_{21}(u) & -q_{21}(u) \end{pmatrix} \tag{16}$$

As mentioned previously, a continuous time Markov chain will allow for a parameter reduction if certain structures are present. Furthermore, identifying such structures will make the model more theoretically tractable. As a simple illustration of such a model, consider the case where there are four states, i.e. $N = 4$. State 1 corresponds to the vehicle being parked at home. State 2 corresponds to the vehicle being on a trip that started from home. State 3 corresponds to the vehicle being parked somewhere else. State 4 corresponds to the vehicle starting a trip from somewhere else than at home. The parameter reduction is thus obtained if it is assumed that the vehicle cannot switch directly from being parked at home to being parked somewhere else, that is from states 1 to 3. Also it would be reasonable to assume that the vehicle does not drive from home to return to home, without an intermediate stop. Under these assumptions, the matrix of transition intensities becomes:

$$\mathbf{Q}(u) = \begin{pmatrix} -q_{12}(u) & q_{12}(u) & 0 & 0 \\ 0 & -q_{23}(u) & q_{23}(u) & 0 \\ 0 & 0 & -q_{34}(u) & q_{34}(u) \\ q_{41}(u) & 0 & q_{43}(u) & -(q_{43}(u) + q_{41}(u)) \end{pmatrix}. \tag{17}$$

The discrete time transition probability matrix can then be found by (15). In this case the number of parameters to be estimated for each time step is reduced from $N \cdot (N-1) = 12$ to 5, by formulating the model in continuous time as opposed to discrete time. The idea behind this specific model is that it can capture whether the vehicle is parked for different lengths of time, depending on the location. Also it can capture whether the vehicle is usually parked at home at night. As the number of states in the model increases, and supposing that certain structures can be identified, the parameter reduction gained by formulating the model in continuous time is increased.

C. Hidden Markov Models

The classical time-varying Markov models only allow for modelling states that are observed. Thus, in this setup, if the data at our disposal is only *driving* and *not driving* we are limited to choosing at two-state classical Markov model for describing the data. Furthermore, another important limiting characteristic of time-varying Markov models, is that the time spent in each state is exponentially distributed, albeit with time-varying intensity. This implies that the time until the next transition out of the current state does not depend on the time spent in said state. For models with few states this may be particularly unrealistic.

To address these two important restrictions in a context of limited we introduce a hidden Markov model, which allows us to estimate states that are not directly observed in the data. In such a way that the actual time spent in each observed state is properly captured. A hidden Markov model is obtained by introducing a new state to the original Markov model. The new state is, however, indistinguishable from one or more of the observed states in the original model. This allows for *non-exponentially* distributed waiting times in each of the observed states, while the Markov assumption is satisfied for the extended model with the hidden states. Specifically the time spent in each observed state is a mixture of exponential distributions. It should be stressed that the same results could be obtained using an ordinary Markov model where the hidden states are actually observed in the data. In short, hidden states are meant to fill the lack of state information. A thorough introduction to hidden Markov models can be found in [19], which also includes R-scripts for parameter estimation.

III. Parameter Reduction via B-Splines

As the number of parameters to be estimated is huge, techniques to reduce this number are needed. One such a technique consists of applying B-splines to approximate the diurnal variation. For a thorough introduction to B-splines as well as other methods for parameter reduction such as smoothing splines and kernels, see [20].

A. B-Splines

To construct a B-spline, first define the knot sequence τ such that

$$\tau_1 \leq \tau_2 \leq \cdots \leq \tau_M. \tag{18}$$

Let this sequence of knots be defined on the interval where we wish to evaluate our spline. In this particular case the knots should be placed somewhere in the interval $[0, 1440]$, that is, over the day.

Denote by $B_{i,m}(x)$ the ith B-spline basis function of order m for the knot sequence τ, where $m < M$. The basis functions are defined recursively as follows:

$$B_{i,1}(x) = \begin{cases} 1 & \text{if } \tau_i \leq x < \tau_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

$$B_{i,m}(x) = \frac{x - \tau_i}{\tau_{i+m-1} - \tau_i} B_{i,m-1}(x) + \frac{\tau_i - x}{\tau_{i+m} - \tau_{i+1}} B_{i,m-1}(x) \tag{20}$$

for $i = 1, \ldots, M-m$. These basis functions are polynomials of order $m-1$ taking values on the interval $[\tau_1, \tau_M]$.

A B-spline curve of degree m is a piecewise polynomial curve defined as follows:

$$S_m(x) = \sum_{i=1}^{M-m} C_i B_{i,m}(x), \tag{21}$$

where C_i, $i = \{1, \ldots, M-m\}$, form the *control polygon*. The $B_{i,m}(x)$ are the B-spline basis functions of order m defined over the knot vector.

As we aim at modeling the diurnal variation in the driving pattern, it is reasonable that the basis splines are periodic. This can be achieved by introducing $2m$ new knots to the existing knots. The new knots are defined as follows:

$$\tau_{1-h} = \tau_{M-h} - (\tau_M - \tau_1) \text{ for } h \in \{1, \ldots, m\} \tag{22}$$

$$\tau_{M+h} = \tau_h + (\tau_M - \tau_1) \text{ for } h \in \{1, \ldots, m\}. \tag{23}$$

More specifically, let the vector containing the new knots be represented by $\tau' = \{\tau_{1-m}, \ldots, \tau_{M+m}\}$. For each B-spline basis function, $m+1$ knots are required, though they may be overlapping. The B-spline basis functions are uniquely defined by the position of the knots. In particular, if the knots are shifted by some constant α, the basis functions will be the same as the original, except that they are shifted by α. If the new knot vector is defined as τ', the basis function defined by the knots $\{\tau_M, \ldots, \tau_{M+m}\}$ will be the same as that defined for the knots $\{\tau_{1-m}, \ldots, \tau_1\}$, except that it is shifted by the interval length $\tau_M - \tau_1$. In this way we can define a basis function that is harmonic in the sense that it is recurrent over different days.

All piecewise polynomial splines of order m defined over the knot vector τ can be constructed from the basis functions defined in (19)-(20). Hence using B-splines does not limit the choice of polynomial splines in any way. Nonetheless, an advantage of using B-splines is that the desired spline can be written as a linear combination of predefined basis functions. This proves useful as a generalized linear model can be applied to estimate the transition probabilities. Traditionally cubic B-splines are used, i.e. $m = 4$, which is also the case here. A motivation for using cubic B-splines is that the spline produced will be of order 4 and furthermore, if $\tau_i \neq \tau_j$ for all $i \neq j$, it will be $\mathcal{C}^2$ everywhere. A function which is $\mathcal{C}^2$ is indistinguishable from a $\mathcal{C}^\infty$ to the human eye. For a further discussion on why to choose cubic splines, see [20].

B. A Generalized Linear Model

To reduce the number of parameters in the model, a B-spline can be fitted to the time-varying transition probabilities $p_{jk}(s)$. There are, however, some issues with this approach. Firstly, there is no guarantee that the fitted B-spline is always in the interval $[0, 1]$, which is a problem as we are modeling probabilities. Secondly, if $\sum_{k=1}^{N} n_{jk}(s) = 0$ for some s, the estimate for $p_{jk}(s)$ given by (6) is undefined. A more refined approach is to use a generalized linear model instead. In the following, such an approach is outlined.

Each day, at a specific minute, a transition from state j to state k either occurs or does not occur. Thus for every s on the diurnal cycle we can consider the number of transitions to be binomially distributed, i.e. $n_{jk}(s) \sim B(z_j(s), p_{jk}(s))$, where the number of Bernoulli trials at s, given by $z_j(s) = \sum_{k=1}^{N} n_{jk}(s)$, are known and the probabilities of success, $p_{jk}(s)$, are unknown. The data can now be analyzed using a *logistic regression*, which is a generalized linear model. The explanatory variables in this model are taken to be the basis functions for the B-spline. The logit transformation of the odds of the unknown binomial probabilities are modeled as linear functions of the basis functions $B_{i,m}(s)$. We model $Y_{jk}(s) = n_{jk}(s)/z_j(s)$ and in particular we are interested in $\mathrm{E}\left[Y_{jk}(s)\right] = p_{jk}(s)$.

Next we elaborate on how the logistic regression works in this particular case. For a general treatment of this problem see ([21]).

We shall use a linear model for a function of p, the *link function*. The canonical link for the binomial distribution is the *logit transformation*,

$$\mathrm{logit}(p) = \log\left(\frac{p}{1-p}\right), \tag{24}$$

which is used as the link function. The resulting transformed means are given by $\eta_{jk}(s)$, which is modeled using a *linear model* with the B-spline basis functions as explanatory variables, i.e.,

$$\begin{aligned}\eta_{jk}(s) &= \log\left(\frac{p_{jk}(s)}{1-p_{jk}(s)}\right) \\ &= C_{jk,1}\cdot B_{1,4}(s) + \ldots + C_{jk,M}\cdot B_{M,4}(s).\end{aligned} \tag{25}$$

The linear prediction of $\eta_{jk}(s)$ is therefore given by

$$\widehat{\eta}_{jk}(s) = \widehat{C}_{jk,1}\cdot B_{1,4}(s) + \ldots + \widehat{C}_{jk,M}\cdot B_{M,4}(s), \tag{26}$$

where the estimates, $\widehat{C}_{jk,1}, \ldots, \widehat{C}_{jk,M}$, are found by the *iteratively reweighted least squares method*. The inverse transformation of the link function in (25), which provides the probability of a transition from state j to state k at time s, is the *logistic function*

$$p_{jk}(s) = \frac{\exp(\eta_{jk}(s))}{1+\exp(\eta_{jk}(s))}. \tag{27}$$

The estimates of the transition probabilities are thus given by

$$\widehat{p}_{jk}(s) = \frac{\exp\left(\widehat{\eta}_{jk}(s)\right)}{1+\exp\left(\widehat{\eta}_{jk}(s)\right)},\ \forall j,k. \tag{28}$$

The procedure of applying a *generalized linear model* is implemented in the statistical software package R as the function `glm(·)`.

C. Choosing the Knots

Choosing the amount and position of the knots in the knot vector τ is important to obtain a good fit for the model. A naive method for placing the knots is to distribute them uniformly over the day. A uniform positioning, however, does not take into account the peakedness of the estimate of $p_{jk}(s)$. An algorithm for placing the knots is given in [22].

The proposed algorithm for placing the knots runs as follows:

1) Decide first on the total number of knots, M.
2) Decide next on an initial number of knots, $M_{init} < M$, to be dispersed uniformly in the interval, with one at each endpoint. Denote these knots by τ_{init}.
3) Fit the model and calculate the likelihood for each knot interval.
4) Find the two adjacent knots with the lowest likelihood of the model in this interval. Denote these knots $\{\tau_j, \tau_{j+1}\}$.
5) Place a new knot, $\tau^\star$, in the middle of the interval (τ_j, τ_{j+1}).
6) Go to step 3 if the new number of knots $M^\star < M$. If $M^\star = M$ then stop.

Once an algorithm for distributing the knots is in place, the number of knots to choose, M, has to be decided. If the number of knots is too low, the model can be improved by placing additional knots. If the amount of knots is too high, the model is overparameterized. We recommend therefore testing different models recursively up to some large M, and choosing the number of knots where there does not seem to be any significant improvement beyond this point.

IV. Numerical Example

In this section the model is fitted to a sample of data collected from the utilization of a single vehicle. The data set solely contains information on whether the vehicle is *driving* or *not driving*. To this avail, we introduce hidden driving states to accurately model the duration of the trips. We let the first state denote the vehicle being parked and the others be various driving states. In the estimation procedure we calculate first the off-diagonal elements of $\mathbf{P}(\mathbf{s})$ and then compute the diagonal elements by $\widehat{p}_{ii}(s) = 1 - \sum_{j=1, j\neq i}^{N} \widehat{p}_{ij}(s)$. Let now $\widehat{p}_{1\cdot}(s) = \sum_{j=1, j\neq i}^{N} \widehat{p}_{ij}(s)$. As we cannot distinguish between the different hidden states, we first determine $\widehat{p}_{1\cdot}(s)$ and then we estimate the probabilities driving the transition into the different hidden states, $\widehat{p}_{1j}(s)$ for $j \in \{2, \ldots, N\}$, as a fixed proportion of $\widehat{p}_{1\cdot}(s)$. Also $p_{1\cdot}(s)$ is the most interesting parameter, as it is the probability of starting a trip within the next minute, conditional on the vehicle not driving at time s.

A. Data

The example is based on GPS-based data pertaining to a single vehicle in Denmark in the period spanning the five months from 31-10-2002 to 29-03-2003, with a total of 150 days. Our aim is to model the use of this vehicle. The data set only shows whether the vehicle was *driving* or *not driving* at any given time. No other information was provided in order to protect the privacy of the vehicle owner. The data

set comprises of a total of 799 trips. The time resolution is in minutes.

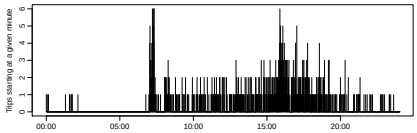

Fig. 1. *Trips starting at a certain minute of the day, cumulated for 107 weekdays.*

The dataset has been split into two main periods, weekdays and weekends. Figure 1 illustrates the number of trips starting at a given minute for the weekdays. Notice that there is a significant degree of diurnal variation, with a lot of trips starting around 07:00 and again around 16:00. Also there are no observations of trips starting between 02:30 and 06:00. Other patterns are found for weekends, but as the approach is similar, we focus in the following on trips starting on weekdays. Driving patterns may also exhibit annual variations, however the limited data sample does not allow for capturing this.

B. Estimation

Firstly, naive B-splines have been fitted to the data using the logistic regression and the result is shown in Figure 2. These B-splines are described as naive in the sense that the knots defining the basis functions for the B-splines are placed uniformly over the 1-day interval. The gray lines are the estimates $\widehat{p}_{1\cdot}(s)$ obtained from (6). As the number of basis-functions increases, it is apparent that the fit improves.

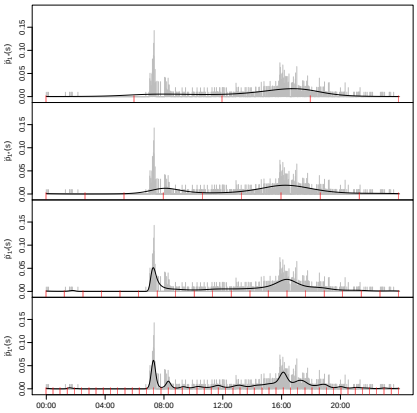

Fig. 2. *From top to bottom: Fitting the estimate $\widehat{p}_{1\cdot}(s)$ where the knots are uniformly distributed in the interval from 00:00 to 23:59 on a weekday, with number of knots $\{5, 10, 20, 50\}$. For reference, the gray bars are the estimates of $\widehat{p}_{1\cdot}(s)$ from (6) with no parameter reduction. The red bars indicate the knot positions.*

The algorithm for placing the knots is implemented considering an initial amount of knots $M_{init} = 7$. The left plot in Figure 3 shows the recursive tests for different number of knots, M_n.

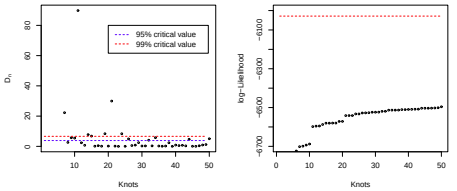

Fig. 3. *Left: Log-likelihood ratio test statistic, given by D_n, from the model with n knots vs. the model with $n-1$ knots. 95% and 99% critical values are shown for a χ^2-distribution with one degree of freedom. Right: The log-likelihood of the models with different knots. The red dashed line is the likelihood of the model with estimates based on (6).*

Referring to Figure 3, left, the model with a total number of knots $M_n = 21$ is chosen, as no significant improvement is attained beyond this point. In Figure 3, right, the log-likelihood for models with different numbers of knots is shown. The red dashed line is the log-likelihood of the model with the estimates found by (6) and corresponds to a perfect data fit. It is in some sense a limit for the fitted models.

The models based on B-splines are sub-models of the model in which a knot is placed at every minute. In this model, the transition probabilities are estimated independently for every minute, and in turn the model corresponds to that with no parameter reduction. The models where the number of parameters is reduced can be tested against the model with no parameter reduction. This leads to a test statistic that will be χ^2-distributed with $1440 - M$ degrees of freedom for each time-varying transition probability. Accordingly the critical value will be very large (> 1475 for estimating one time-varying transition probability for $M \leq 50$ at 95% significance) and thus a test for sufficiency is not appropriate, as it is difficult to test anything.

The top plot in Figure 4 illustrates the estimate of $\widehat{p}_{1\cdot}(s)$ using B-splines with $M = 21$, where the knots are placed by the algorithm introduced in Section 3.3. For comparison, the model with the naive knots and $M = 21$ is shown on the bottom plot in Figure 4. By visual inspection, it is observed that the model in which the knots are placed according to the algorithm in Section 3.3 better captures the peakedness of $p_{1\cdot}(s)$.

From the estimation of the transition probabilities and the evaluation of likelihoods we conclude that a Markov model including five hidden time-homogeneous driving states and one inhomogeneous non-driving state satisfactorily describes the use of the vehicle. Indeed, with such a number of states, we manage to properly capture the trip lengths. In Figure 5 the empirical trip lengths obtained from the data are plotted along with the distribution of the trip lengths based on the fitted model. Notice that the distribution is clearly not exponential.

C. Applications

The applications of the proposed stochastic model for driving patterns range from simulating different driving scenarios

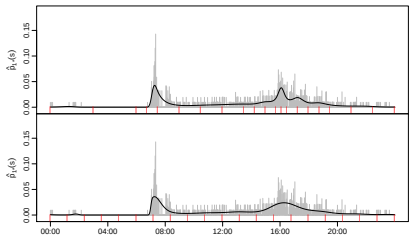

Fig. 4. *Top: $\widehat{p}_{1\cdot}(s)$ based on the B-splines with $M = 21$ and the knots placed using the algorithm, plotted as the black line over the estimates $\widehat{p}_{1\cdot}(s)$ with no parameter reduction. Bottom: $\widehat{p}_{1\cdot}(s)$ based on the naive B-splines with $M = 21$, plotted as the black line. The red bars indicate the knot position.*

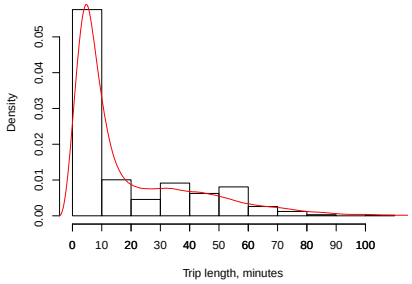

Fig. 5. *Trip lengths, empirical in histogram bars, and estimated density, in red, based on Monte Carlo simulation from fitted model.*

to calculating the probability of a trip starting within a given interval. In addition, the model is prerequisite to determine the optimal charging scheme for an electric vehicle.

1) Probabilities and Simulations: Four driving scenarios are simulated and shown in Figure 6. Markov states are indicated in a binary form depending on whether the vehicle is driving "1" or not "0".

Next we illustrate how to find the probability of a trip starting within a given interval. Suppose that at time s the vehicle is parked. Denote the waiting time until the next trip starts by Z_s. We have that $Z_s \sim \exp(q_{1\cdot}(s))$, where $q_{1\cdot}(s) = \sum_{j=2}^{N} q_{1j}(s)$. The probability of a trip starting within the time interval $[s, s+\tau]$ is thus

$$\mathbb{P}(Z_s \leq \tau) = 1 - e^{-\int_0^\tau q_{1\cdot}(s+t)dt}. \tag{29}$$

Using this equation, for example, the probability of a trip starting in the interval from 00:00 to 06:00 is $\mathbb{P}(Z_{00:00} \leq 06{:}00) = 1 - e^{-\int_0^{6\cdot 60} q_{1\cdot}(t)dt} = 0.098$.

In the top part of Figure 7, the probability of starting a trip within the next hour, conditional on not driving at the beginning of that hour, is depicted. The probability of the vehicle being in use at any time of the day is found using bootstrap ([23]) and is shown in the bottom part of Figure 7.

2) Example of Electric Vehicle Charging: A stylized example of charging an EV is considered below. Suppose that the owner desires to have the vehicle fully charged for the next time he/she must use the vehicle. At the same time, he/she wishes to minimize the cost of doing so. As the use of the vehicle is stochastic, it can only be guaranteed that the vehicle is fully charged with some probability. Suppose that the vehicle is parked at 21:00 hours and that the battery is at half capacity. Furthermore, we assume that the battery has a total capacity of 20 kWh and that it takes 4 hours to fully charge the battery, if it is completely empty. In addition we assume a stylized price signal, according to which the electricity price is 60 EUR/MWh until 00:00 hours, and then drops to 30 EUR/MWh. We consider that the driving behavior is independent of the electricity price.

In Figure 8 the different charging strategies are presented for varying levels of certainty. It can be observed that lowering the desired level of certainty will delay the charging. As a consequence, the cost of charging is also reduced, which is seen in Table I.

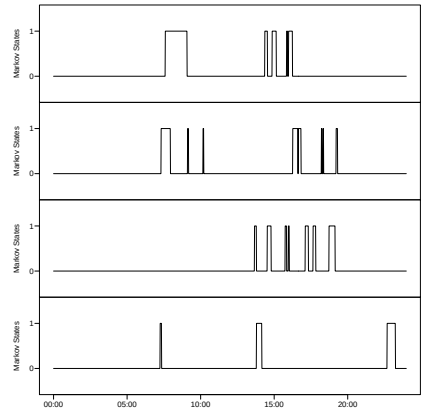

Fig. 6. *Four distinct realizations of driving patterns using the proposed stochastic model.*

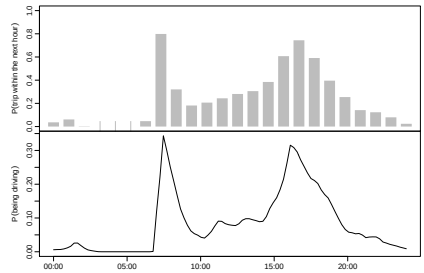

Fig. 7. *Top: A graph plotting the probability of starting a trip within the next hour, conditional on not driving at the present time, which is found by applying (29). Bottom: The probability of the vehicle being in use at any time of the day, which is estimated using bootstrap.*

Certainty	95%	90%	75%	50%
Cost (EUR)	NA	0.60	0.45	0.30

TABLE I
THE ASSOCIATED COSTS FOR DIFFERENT LEVELS OF CERTAINTY.

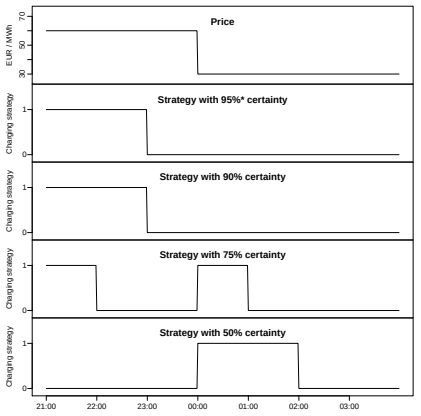

Fig. 8. *Top: The electricity price. Bottom four: The charging strategies with decreasing certainty of the vehicle being fully charged for the next trip. A "1" indicates that the vehicle is being charged and a "0" indicates that the vehicle is not charging. * having 95% certainty is not feasible, see Table I.*

V. Conclusion and Future Research

This paper proposes a suitable model that captures the diurnal variation in the use of a vehicle. The number of parameters is significantly reduced by using B-spline basis functions as explanatory variables in a logistic regression. The model is versatile and can be applied to describe driving data from any single vehicle, thus providing a reliable model for the use of that vehicle.

It would be interesting to apply the model to data that includes location, to see how this affects the model. The model could be extended to cover a population of vehicles by using a mixed-effect model. Another extension to the model could be to estimate the transition probabilities adaptively in time. This way structural changes in the driving behaviour of the vehicle user, such as variation over the year or a change in use as a result of from the household purchasing an additional vehicle, could be captured. An obvious next step is to use the model to implement a charging strategy that minimizes the costs of driving considering the underlying uncertainty in the use of the vehicle.

References

[1] T. Golob and J. Gould, "Projecting use of electric vehicles from household vehicle trials," *Transportation Research Part B: Methodological*, vol. 32, no. 7, pp. 441–454, 1998.

[2] N. Pearre, W. Kempton, R. Guensler, and V. Elango, "Electric vehicles: How much range is required for a day's driving?" *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1171–1184, 2011.

[3] R. Green, L. Wang, and M. Alam, "The impact of plug-in hybrid electric vehicles on distribution networks: a review and outlook," in *Power and Energy Society General Meeting, 2010 IEEE*, july 2010, pp. 1 –8.

[4] N. Rotering and M. Ilic, "Optimal charge control of plug-in hybrid electric vehicles in deregulated electricity markets," *Power Systems, IEEE Transactions on*, vol. 26, no. 3, pp. 1021 –1029, aug. 2011.

[5] S. Acha, T. C. Green, and N. Shah, "Effects of optimised plug-in hybrid vehicle charging strategies on electric distribution network losses," in *Transmission and Distribution Conference and Exposition, 2010 IEEE PES*, april 2010, pp. 1 –6.

[6] L. Gan, U. Topcu, and S. Low, "Optimal decentralized protocol for electric vehicle charging," California Institute of Technology, Tech. Rep., 2011. [Online]. Available: http://www.cds.caltech.edu/ utopcu/images//9/96/GTL-cdc11.pdf

[7] A. Lojowska, D. Kurowicka, G. Papaefthymiou, and L. van der Sluis, "From transportation patterns to power demand: Stochastic modeling of uncontrolled domestic charging of electric vehicles," in *Power and Energy Society General Meeting, 2011 IEEE*, july 2011, pp. 1 –7.

[8] Z. Ma, D. Callaway, and I. Hiskens, "Decentralized charging control for large populations of plug-in electric vehicles: Application of the Nash certainty equivalence principle," in *Control Applications (CCA), 2010 IEEE International Conference on*, sept. 2010, pp. 191 –195.

[9] P. Richardson, D. Flynn, and A. Keane, "Optimal charging of electric vehicles in low-voltage distribution systems," *Power Systems, IEEE Transactions on*, vol. 27, no. 1, pp. 268 –279, feb. 2012.

[10] R. Bessa, F. Soares, J. Pecas Lopes, and M. A. Matos, "Models for the ev aggregation agent business," in *PowerTech, 2011 IEEE Trondheim*. IEEE, 2011, pp. 1–8.

[11] P. Grahn, K. Alvehag, and L. Soder, "Plug-in-vehicle mobility and charging flexibility Markov model based on driving behavior," in *European Energy Market (EEM), 2012 9th International Conference on the*. IEEE, 2012, pp. 1–8.

[12] G. Hill, P. Blythe, and C. Higgins, "Deviations in Markov chain modeled electric vehicle charging patterns from real world data," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*. IEEE, 2012, pp. 1072–1077.

[13] J. Rolink and C. Rehtanz, "Estimation of the availability of grid-connected electric vehicles by non-homogeneous semi-Markov processes," in *PowerTech, 2011 IEEE Trondheim*. IEEE, 2011, pp. 1–7.

[14] F. Soares, J. P. Lopes, P. R. Almeida, C. Moreira, and L. Seca, "A stochastic model to simulate electric vehicles motion and quantify the energy required from the grid," in *17th Power Systems Computation Conference*, 2011, pp. 22–26.

[15] G. Souffran, L. Miegeville, and P. Guerin, "Simulation of real-world vehicle missions using a stochastic Markov model for optimal design purposes," in *Vehicle Power and Propulsion Conference (VPPC), 2011 IEEE*. IEEE, 2011, pp. 1–6.

[16] A. Schlote, E. Crisostomi, S. Kirkland, and R. Shorten, "Traffic modelling framework for electric vehicles," *International Journal of Control*, vol. 85, no. 7, pp. 880–897, 2012.

[17] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, ser. Texts from Oxford University Press. Oxford University Press, 2001. [Online]. Available: http://books.google.dk/books?id=G3ig-0M4wSIC

[18] Y. Pawitan, *In All Likelihood: Statistical Modelling and Inference using Likelihood*, ser. Oxford science publications. Clarendon Press, 2001. [Online]. Available: http://books.google.dk/books?id=M-3pSCVxV5oC

[19] W. Zucchini and I. MacDonald, *Hidden Markov Models for Time Series: An Introduction Using R*, ser. Monographs on Statistics and Applied Probability. Taylor & Francis Group, 2009. [Online]. Available: http://books.google.dk/books?id=LDDzvCsdVs8C

[20] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, ser. Springer series in statistics. Springer, 2008. [Online]. Available: http://books.google.dk/books?id=tVIjmNS3Ob8C

[21] H. Madsen and P. Thyregod, *Introduction to General and Generalized Linear Models*, ser. Chapman & Hall/CRC Texts in Statistical Science. Chapman & Hall/CRC, 2010. [Online]. Available: http://books.google.dk/books?id=JhuDNgAACAAJ

[22] W. Mao and L. H. Zhao, "Free-knot polynomial splines with confidence intervals," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 65, no. 4, pp. 901–919, 2003. [Online]. Available: http://dx.doi.org/10.1046/j.1369-7412.2003.00422.x

[23] A. Davison and D. Hinkley, *Bootstrap Methods and Their Application*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997. [Online]. Available: http://books.google.dk/books?id=4aCDbm_t8jUC

Paper B

Optimal Charging of an Electric Vehicle using a Markov Decision Process

Published in *Applied Energy, 2014*.

Applied Energy 123 (2014) 1–12

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Optimal charging of an electric vehicle using a Markov decision process

Emil B. Iversen *, Juan M. Morales, Henrik Madsen

DTU Compute, Technical University of Denmark, Matematiktorvet, Building 322, DK-2800 Lyngby, Denmark

HIGHLIGHTS

- This paper proposes an algorithm to optimally charge an electric vehicle considering the usage of the vehicle.
- The charging policy depends on the use of the vehicle, the risk aversion of the end-user, and the electricity price.
- The model is versatile and can easily be adapted to any specific vehicle, thus providing a customized charging policy.

ARTICLE INFO

Article history:
Received 5 October 2013
Received in revised form 4 February 2014
Accepted 4 February 2014
Available online 12 March 2014

Keywords:
Electric vehicles
Driving patterns
Optimal charging
Markov processes
Stochastic dynamic programming

ABSTRACT

The combination of electric vehicles and renewable energy is taking shape as a potential driver for a future free of fossil fuels. However, the efficient management of the electric vehicle fleet is not exempt from challenges. It calls for the involvement of all actors directly or indirectly related to the energy and transportation sectors, ranging from governments, automakers and transmission system operators, to the ultimate beneficiary of the change: the end-user. An electric vehicle is primarily to be used to satisfy driving needs, and accordingly charging policies must be designed primarily for this purpose. The charging models presented in the technical literature, however, overlook the stochastic nature of driving patterns. Here we introduce an efficient stochastic dynamic programming model to optimally charge an electric vehicle while accounting for the uncertainty inherent to its use. With this aim in mind, driving patterns are described by an inhomogeneous Markov model that is fitted using data collected from the utilization of an electric vehicle. We show that the randomness intrinsic to driving needs has a substantial impact on the charging strategy to be implemented.

1. Introduction

Electric vehicles (EVs) are emerging as a sustainable and environmentally friendly alternative to conventional vehicles, provided that the energy used for their charging is obtained from renewable energy sources. The energy generated from renewable sources such as sunlight, wind and waves is, however, dependent on weather conditions. As a consequence, the electricity production from these sources is inherently uncertain in time and quantity. Furthermore, electricity has to be produced and consumed at the same time, as the large-scale storage of the energy generated is, still today, very limited. As a result, the energy obtained from renewables may be wasted in times when the demand for electricity is not high enough to absorb it, with a consequent detrimental effect on the profitability of renewables. Since the battery in an EV is basically a storage device for energy, the large-scale integration of EVs in the transportation sector may contribute to substantially increasing the socioeconomic value of an energy system with a large renewable component, while reducing the dependence of the transportation sector on liquid fossil fuel.

For this reason, EVs have received increased interest from the scientific community in recent years (detailed literature reviews of the state of the art can be found in [1,2]). Special attention has been given to the analysis of the effect of EVs integration on the electricity demand profile [3,4], emissions [5] and social welfare [6–8], and to the design of charging schemes that avoid increasing the peak consumption [9,10], help mitigate voltage fluctuations and overload of network components in distribution grids [11], and/or get the maximum economic benefit from the storage capability of EVs within a market environment, either from the perspective of a single vehicle [12,13] or the viewpoint of an aggregator of EVs [14,15]. In all these publications, though, and more generally in the technical literature on the topic, the charging problem of an EV is addressed either by considering deterministic driving patterns, when the focus is placed on the management of a single vehicle, or by aggregating the driving needs of different EV

* Corresponding author. Tel.: +45 45 25 30 75.
E-mail addresses: jebi@dtu.dk (E.B. Iversen), jmmgo@dtu.dk (J.M. Morales), hm@dtu.dk (H. Madsen).

http://dx.doi.org/10.1016/j.apenergy.2014.02.003
0306-2619/

users, when the emphasis is on modeling a whole fleet of EVs. This aggregation, however, obscures the dynamics of each specific vehicle. Likewise, the deterministic driving patterns of a single EV are often based on expected values or stylized behaviors, which fail to capture important features of the charging problem such as the daily variation in the use of the vehicle or potential user conflicts in terms of not having the vehicle charged and ready for use. A stochastic model for driving patterns provides more insight into these aspects and becomes fundamental for applying a charging scheme in the real world. Despite this, the stochastic modeling of driving patterns has received little attention from the scientific community, as pointed out in [1]. We mention here the research work by [16], in which they aim to capture the uncertainty intrinsic to the vehicle use by means of a Monte Carlo simulation approach. They assume, however, an uncontrolled charging scheme.

The work developed in this paper departs from the following two premises:

1. The primary purpose of the battery of an EV is to provide power to drive the vehicle and not to store energy from the electricity grid. Consequently, it is essential that enough energy is kept in the battery to cover any desired trip. This calls for a decision tool that takes into account the driving needs of the EV user to determine when charging can be postponed and when the battery should be charged right away.
2. The complexity of human behavior points to a stochastic model for describing the use of the vehicle. In turn, this stochastic model should be integrated into the aforementioned decision tool and exploited by it.

That being so, this paper introduces an algorithm to optimally decide when to charge an EV that exhibits a stochastic driving pattern. The algorithm builds on the inhomogeneous Markov model proposed in [17] for describing the stochastic use of a single vehicle. The model parameters are then estimated on the basis of data from the use of the specific vehicle. The approach captures the diurnal variation of the driving pattern and relies only on the assumption that the EV-user's driving habits can be explained and modeled as a stochastic process, more particularly, as an inhomogeneous Markov chain. This makes our modeling approach noticeably general and versatile. Our algorithm thus embodies a *Markov decision process* which is solved recursively using a stochastic dynamic programming approach. The resulting decision-support tool allows for addressing issues related to charging, vehicle-to-grid (V2G) schemes [12,18], availability and costs of using the vehicle. The algorithm runs swiftly on a personal computer, which makes it feasible to implement on an actual EV.

The remainder of this paper is organized as follows: In Section 2 the stochastic model for driving patterns developed in [17] is briefly described, tailored to be used in the present work, and extended to address the problem of driving data limitations through hidden Markov models. Section 3 introduces the algorithm for the optimal charging of an EV as a Markov decision process that is solved using stochastic dynamic programming. Section 4 provides results from a realistic case study and explores the potential benefit of implementing V2G schemes. Section 5 concludes and provides directions for future research within this topic.

2. A stochastic model for driving patterns

In this section we summarize and extend the stochastic model for driving patterns developed in [17]. We refer the interested reader to this work for a detailed description of the modeling approach.

2.1. Standard Markov model

A state-space model is considered to describe the use of the EV. In its simplest form, it contains two states, according to which the vehicle is either *driving* or *not driving*. A more extensive version of the model would include a larger number of states which could capture information about where the vehicle is parked, how fast it is driving or what type of trip it is on. The basics of the general multi-state stochastic model are described in this section, including how to fit a specific model on an observed data set.

Let X_t, where $t \in \{0, 1, 2, \ldots\}$, be a sequence of random variables that takes on values in the countable set S, called the state space. Denote this sequence as X. We assume a finite number, N, of states in the state space. A Markov chain is a random process where future states, conditioned on the present state, do not depend on the past states [19]. In discrete time X is a Markov chain if

$$\mathbb{P}(X_{t+1} = k | X_0 = x_0, \ldots, X_t = x_t) = \mathbb{P}(X_{t+1} = k | X_t = x_t) \tag{1}$$

for all $t \geqslant 0$ and all $\{k, x_0, \ldots, x_t\} \in S$.

A Markov chain is uniquely characterized by the transition probabilities, $p_{jk}(t)$, i.e.

$$p_{jk}(t) = \mathbb{P}(X_{t+1} = k | X_t = j). \tag{2}$$

If the transition probabilities do not depend on t, the process is called a homogeneous Markov chain. If the transition probabilities depend on t, the process is known as an inhomogeneous Markov chain.

When it comes to the use of a vehicle, it is appropriate to assume that the probability of a transition from state j to state k is similar on specific days of the week. Thus, for instance, Thursdays in different weeks will have the same transition probabilities. For convenience we further assume that all weekdays (Monday through Friday) have the same transition probabilities. In other words, we consider that the transition probabilities of the inhomogeneous Markov chain vary within the day, but not from day to day. These assumptions can be easily relaxed or interchanged with other assumptions and as such, are not essential to the model. With a sampling time in minutes, and taking into account that there are 1440 min in a day, this leads to the assumption:

$$p_{jk}(t) = p_{jk}(t + 1440). \tag{3}$$

This assumption implies that the transition probabilities, defined by (2), are constrained to be a function of the time, s, in the diurnal cycle. Let the matrix containing the transition probabilities be denoted by $\mathbf{P}(s)$. This matrix characterizes the driving pattern of the specific vehicle under consideration using N states. It has the form:

$$\mathbf{P}(s) = \begin{pmatrix} p_{11}(s) & p_{12}(s) & \cdots & p_{1N}(s) \\ p_{21}(s) & p_{22}(s) & \cdots & p_{2N}(s) \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1}(s) & p_{N2}(s) & \cdots & p_{NN}(s) \end{pmatrix}, \tag{4}$$

where $p_{jj}(s) = 1 - \sum_{i=1, i \neq j}^{N} p_{ji}$.

Now let $n_{jk}(s)$ define the number of observed transitions from state j to state k at time s. From the conditional likelihood function, the maximum likelihood estimate of $p_{jk}(s)$ for the inhomogeneous Markov chain can be found as:

$$\hat{p}_{jk}(s) = \frac{n_{jk}(s)}{\sum_{k=1}^{N} n_{jk}(s)}. \tag{5}$$

A discrete time Markov model can be formulated based on the estimates of $\mathbf{P}(1), \mathbf{P}(2), \ldots, \mathbf{P}(1440)$. One apparent disadvantage of such a discrete time model is its huge number of parameters, namely $N \times (N-1) \times 1440$, where $N \times (N-1)$ parameters have

to be estimated for each time step. Needless to say, the number of parameters to be estimated increases as the number of states grows. We refer to [17] for further details on techniques to reduce the number of parameters to be estimated for each time step for models with more than two states. Another problem is linked to the number of observations available to properly carry out the estimation, i.e. if $\sum_{k=1}^{N} n_{jk}(s') = 0$ for some s', then $\hat{p}_{jk}(s')$ is undefined.

To deal with the large number of parameters as well as undefined transition probability estimates, B-splines are applied to capture the diurnal variation in the driving pattern through a *generalized linear model*. The procedure of applying a *generalized linear model* is implemented in the statistical software package R as the function `glm (·)`. For a thorough introduction to B-splines see [20] and for a general treatment of generalized linear models see [21]. Next we elaborate on how the fitting of the Markov chain model works in our particular case.

Each day, at a specific minute, a transition from state j to state k either occurs or does not occur. Thus for every s on the diurnal cycle we can consider the number of transitions to be binomially distributed, i.e. $n_{jk}(s) \sim B(z_j(s), p_{jk}(s))$, where the number of Bernoulli trials at s, given by $z_j(s) = \sum_{k=1}^{N} n_{jk}(s)$, is known and the probability of success, $p_{jk}(s)$, is unknown. The data can now be analyzed using a *logistic regression*, which is a generalized linear model [21]. The explanatory variables in this model are taken to be the basis functions for the B-spline. The logit transformation of the odds of the unknown binomial probabilities are modeled as linear combinations of the basis functions. We model $Y_{jk}(s) = n_{jk}(s)/z_j(s)$ and in particular, we are interested in $\mathrm{E}[Y_{jk}(s)] = p_{jk}(s)$.

As the basis functions for the B-spline are uniquely determined by the knot vector τ, deciding the knot position and the amount of knots is important to obtain a good fit for the model. Here we proceed as follows: First a number of knots are placed on the interval $[0, 1440]$, with one at each endpoint and equal spacing between them. Denote this initial vector of knots by τ_{init}. The model is then fitted using the basis functions as explanatory variables. Next, the fit of the model between the knots is evaluated via the likelihood function and an additional knot is placed in the center of the interval with the lowest likelihood value. The new knot vector is then given by τ'. We repeat this procedure until the desired number of knots is reached. To determine the appropriate number of knots and avoid over-parametrization, on the basis of a likelihood ratio principle, we test that adding a new knot does significantly improve the fit.

2.2. Hidden Markov models

Standard Markov models can only include states that are explicitly recorded in the data. Thus, if the data only provides information on whether the vehicle is either *driving* or *not driving*, the standard Markov model is restricted to having two states: *driving* or *not driving*. Standard Markov models also result, by default, in the time spent in each state being exponentially distributed, although it may be with time-varying intensity. Accordingly, in a standard Markov model, the time until a transition from the current state to another does not depend on the amount of time already spent in the current state. In the case of a vehicle, this implies that the probability of ending a trip does not depend on the duration of the trip so far. This seems unrealistic for a model capturing the actual use of a vehicle.

To overcome these limitations, we can use a hidden Markov model, which allows estimation of additional states that are not directly observed in the data. In fact, we can estimate these states so that the waiting time in each state matches that which is actually observed in the data. Adding a hidden state is done by introducing a new state in the underlying Markov chain. The new state, however, is indistinguishable from any of the previously observed states. This allows for the waiting time in each observable state to be the sum of exponential variables, which is a more versatile class of distributions. It is worth insisting that the use of hidden Markov models is justified here to address insufficient state information in our data, which only include whether the vehicle is *driving* or *not driving*. Indeed, the same results could be obtained using the underlying Markov chain without hidden states, provided that the hidden states could be observed. In practice, though, more detailed driving data (e.g. including driving speed and/or location of the vehicle) could be available once the actual implementation is made on a vehicle, which in turn would avert the need for a hidden Markov model. For a detailed introduction to hidden Markov models, see [22], where techniques and scripts for estimating parameters are also provided.

The hidden Markov model consists of two parts. First, an underlying unobserved Markov process, $\{X_t : t = 1, 2, \ldots\}$, which describes the actual state of the vehicle. This part corresponds to the Markov model with no hidden states as described previously. The second part of the model is a state-dependent process, $\{Z_t : t = 1, 2, \ldots\}$, such that when X_t is known, the distribution of Z_t depends only on the current state X_t. A hidden Markov model is thus defined by the state-dependent transition probabilities, $p_{jk}(t)$, as defined for the standard Markov chain and the state-dependent distributions given by (in the discrete case):

$$d_{zk}(t) = \mathbb{P}(Z_t = z | X_t = k). \tag{6}$$

Collecting the $d_{zk}(t)$'s in the matrix $\mathbf{D}(z_t)$, the likelihood of the hidden Markov model is given by:

$$L_T = \delta \mathbf{D}(z_1) P(2) \mathbf{D}(z_2), \ldots, P(T) \mathbf{D}(z_T), \tag{7}$$

where δ is the initial distribution of X_1. We can now maximize the likelihood of observations to find the estimates of the transition probabilities between the different hidden states.

2.3. Fitting the Data

The data at our disposal is from the utilization of a single vehicle in Denmark in the period spanning the six months from 23-10-2002 to 24-04-2003, with a total of 183 days. The data is GPS-based and follows specific cars. One car has been chosen and the model is intended to describe the use of this vehicle accordingly. The data set only contains information on whether the vehicle was *driving* or *not driving* at any given time. No other information was provided in order to protect the privacy of the vehicle owner. The data is divided into two periods, a training period for fitting the model from 23-10-2002 to 23-01-2003, and a test period from 24-01-2003 to 24-04-2003 for evaluating the performance of the model. The data set consists of a total of 749 trips. The time resolution is in minutes.

We shall consider a model with one *not driving* state and two (hidden) *driving* states. In other words, one can observe whether the vehicle is driving, but cannot identify *which type of driving state* the vehicle is in. Besides, the hidden *driving* states are not directly interpretable from the data. In practice, they could correspond to driving in different environments (urban/rural) or at different speeds. Be as it may, the inclusion of the hidden structure allows for the probability of ending the current trip to depend on the time since departure, as the vehicle may pass through different driving states before ending the trip. We then compute the transition probability between the hidden states in such a way that the resulting probability distribution of the trip duration follows the one reflected in the data. Furthermore, to fit the model to the data, we assume that only the transition probability from the *not driving* state depends on the time of day. This is done to reduce the

complexity of the estimation procedure, as it is cumbersome to estimate the time-varying parameters of a hidden Markov model.

We now elaborate on the fitting of the hidden Markov model, which is split into estimation of its time-varying and time-invariant parameters.

2.3.1. Fitting time-varying parameters

We need to estimate the probability of a transition from the vehicle being parked to a driving state. We denote this transition estimate by $\hat{p}_{1\cdot}(s)$. It holds that $\hat{p}_{1\cdot}(s) = 1 - \hat{p}_{11}(s)$. Since both the parked state and the transitions from it are directly observable in the data, we can use the procedure described in Section 2.1 to estimate $\hat{p}_{1\cdot}(s)$.

The data have been divided into two main periods: weekdays and weekends. The observed number of trips starting every minute for the weekdays is displayed in Fig. 1. A high degree of diurnal variation is found, with a lot of trips starting around 06:00 and again around 16:00. Also, there are no observations of trips starting between 00:00 and 05:00. Other patterns are found for weekends, but as these do not involve any methodological difference, we limit ourselves to trips starting on weekdays. Annual variations may also be present, however the limited data sample does not allow for capturing such seasonality.

The plot in Fig. 2 illustrates the estimate of $\hat{p}_{1\cdot}(s)$ using B-splines with eight initial knots placed uniformly on the interval and 22 knots in total.

2.3.2. Fitting time-invariant parameters

The time-invariant parameters are to be estimated so that an appropriate probability distribution is fitted to the duration of the trips. The time-invariant parameters are estimated by maximizing the likelihood given in (7). For a given number of *driving* states, the transition probabilities can be estimated using the approach in [22]. Once a model with N states is fitted, we can test if adding an additional state significantly improves the fit. As a model with N states is a sub-model of one with $N+1$ or more states, we increase the number of states until no significant improvement test is observed according to the likelihood ratio.

Fig. 3 represents the histogram of the empirically observed trip lengths along with the theoretical density function of the trip lengths obtained from the fitted model. We use a model with two driving states, as no significant improvement is found beyond this number. Notice that the distribution of the empirically observed trip lengths is adequately captured by the hidden Markov model, although the number of observed trips in the range from 10 to 20 min has a higher prevalence than the fitted distribution. In practical applications, more information could be available to model the behavior of the vehicle (e.g. its location and speed), which should facilitate the modeling of the driving patterns.

In the following section, the algorithm for optimally charging the EV is presented. The optimization algorithm makes use of the transition probabilities characterizing the stochastic model for the driving patterns. Thus, the optimization algorithm is designed to handle the stochastic nature of the driving needs.

3. A stochastic dynamic programming problem

The problem of charging an EV can be posed as a conflict between two opposing objectives. The end-user desires to have the vehicle charged and ready for use at his/her discretion, while also minimizing the costs of running the vehicle. Demand for electricity varies over the day and so does the electricity generated from renewable sources. This introduces a varying energy price which can make it beneficial for the end-user to postpone charging his/her vehicle. This means the user is faced with the problem of postponing charging to minimize costs or to charge right away so as to maximize the availability of the vehicle.

Only the procurement cost of the electricity that is needed to charge, and thus run, the vehicle is considered. That is, we do not take into account investment or maintenance costs. However, we also add a penalty term to the electricity procurement cost that is used (by the EV owner) to control the level of availability of the EV (the higher this penalty, the lower the probability of running out of battery during a trip). Conceptually, the penalty term can be understood as the economic value the EV owner places on being able to complete a plausible trip. Finally, note that the electricity procurement cost can be negative under a V2G scheme and that maintenance costs (such as those related to battery wear) may have a non-trivial effect on the optimal charging strategy. The analysis and valuation of this effect, however, is left for future work.

The algorithm for optimal charging of the EV is formulated as a stochastic dynamic programming problem. We first define the relevant parameters and variables, and then the state-transition and objective function.

3.1. Parameters

$u_{\max}$	maximum rate of charge of the battery (kW)
$u_{\min}$	minimum rate of charge of the battery (kW)
$e_{\max}$	maximum storage level of the battery (kW h)
$e_{\min}$	minimum storage level of the battery (kW h)
λ_t	time varying electricity price (€/MW h)
ϕ	penalty for violating (unserved) driving needs (€/h)
η_c	charging efficiency of the battery
η_d	discharging efficiency of the battery
v_i	average speed when the vehicle is in use in state i (km/h)
μ_i	drive efficiency in state i (kW h/km)
κ	battery capacity (kW h)
ω	conversion factor from minutes to hours i.e. $\omega = (60)^{-1}$ (h/min)
β	time discount factor

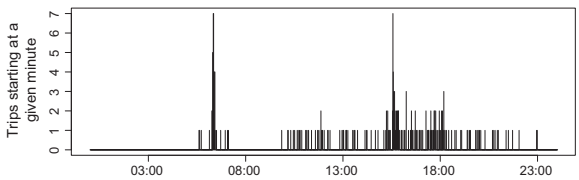

Fig. 1. Number of trips starting at a certain minute of the day, cumulated for the first 66 weekdays.

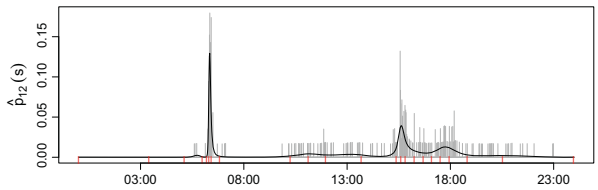

Fig. 2. $\hat{p}_1(s)$ based on the B-splines and the logistic regression, plotted as the black line over the estimates $\hat{p}_1(s)$ from (5), in gray. The red bars indicate the knot positioning. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

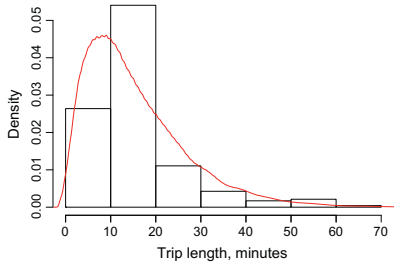

Fig. 3. The empirical distribution of the trip lengths, shown as the histogram bars, and the theoretical density from the fitted model, shown in red, obtained via Monte Carlo simulation and the subsequent kernel density estimation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The maximum rate of charge, $u_{\max}$, reflects a power limit on the electric sockets in a residential household or a technical constraint due to thermal limits on the battery (as batteries generate heat when charged). The minimum rate of charge on the battery, $u_{\min}$, reflects that the battery may be limited to only charging i.e. $u_{\min} = 0$ or that discharging the battery is allowed so as to inject power into the grid, i.e. $u_{\min} < 0$. The bounds on the storage limits on the battery, $e_{\max}$ and $e_{\min}$, reflect the storage capacity of the battery. These limits can also be altered to restrict life-cycle degradation of the battery. The penalty ϕ is the *inconvenience cost* incurred if the vehicle cannot comply with the driving needs. As seen later, this penalty determines the trade-off between the electricity procurement cost and the availability of the EV to cover a plausible trip. The ultimate aim of the penalty is to model the degree of flexibility of the EV user (the higher the penalty, the less flexible the EV user is). The penalty cost is expressed in €/h, indicating how much the EV-user values (in monetary terms) having to postpone a trip one hour because the EV is not available for driving. Therefore, the value of this penalty cost should be set by the EV-user in accordance with his/her economy, driving needs, and attitude towards risk. Parameters η_c and η_d represent the efficiency losses from battery charging and discharging, respectively. The constant v_i is the average speed of the vehicle, when the vehicle is in state i, keeping in mind that the modeling framework is general enough to capture multiple different driving states, say urban and rural. The driving efficiency, μ_i, captures the performance of the vehicle in driving state i. The constant κ is the total energy capacity of the battery. The parameter ω is used as a conversion factor from hours to minutes, as the model inputs are in hourly values and the model is run in 1-min time steps.

3.2. State variables

e_t	total energy stored in the battery at the beginning of minute t (kW h)
x_t	desired driving state, where $x_t \in \{1 \dots N\}$

We assume that variable x_t is exogenously given by the inhomogeneous Markov model described in Section 2. Variable e_t is the energy stored in the battery. We define a state variable at time t as $S_t = (e_t, x_t)$. Notice that, as the driving state is exogenously given, it does not depend on e_t and thus the vehicle is allowed to be in a driving state even though there is no energy on the battery. Logically this is not feasible. Consequently, we refer to x_t as the *desired* driving state, since it can only be reached if there is enough charge on the battery. To cope with this issue, we first define the set, S_D, as the collection of states that x_t can take where the vehicle is driving. Then we define the auxiliary variable x_t^a as the actual driving state, i.e.,

$$x_t^a = \begin{cases} 1 & \text{if } e_t = e_{\min} \wedge x_t \in S_D \\ x_t & \text{else.} \end{cases} \tag{8}$$

Notice that x_t and x_t^a differ only when there is not enough charge to complete the desired trip. State 1 denotes the parked state. Therefore, according to (8) the vehicle is forced to stop when there is not enough charge on the battery to drive any further. Note that S_t implicitly includes x_t^a as a state, inasmuch as x_t^a is derived from x_t and e_t.

3.3. Decision (action) variables

u_t	desired energy charged into (or discharged from) the battery in minute t (kW)

As for the driving state, we define an auxiliary charging variable u_t^a, which is the actual energy charged into the battery, since the vehicle is unable to charge when it is in use. The new variable u_t^a is then defined as follows:

$$u_t^a = \begin{cases} 0 & \text{if } e_t > e_{\min} \wedge x_t \in S_D \\ u_t & \text{else.} \end{cases} \tag{9}$$

Thus u_t^a is zero when the vehicle is actually driving, and equal to u_t otherwise. Again, if both the state S_t and the desired energy charged u_t are known, the actual energy charged u_t^a follows implicitly from these.

3.4. State transition function

The driving state variable x_t evolves randomly according to the inhomogeneous Markov model described in Section 2. The state-transition function for the storage level of the battery can be expressed as:

$$e_{t+1} = e_t + \left(\eta_c \mathbb{1}_{\{u_t^a \geq 0\}} + \frac{1}{\eta_d}\mathbb{1}_{\{u_t^a < 0\}}\right)\omega u_t^a - \sum_{i=1}^{N} \mathbb{1}_{\{x_t^a = i\}} v_i \mu_i \omega. \quad (10)$$

Eq. (10) describes the dynamics of the energy stored in the battery. It defines the storage level at time $t+1, e_{t+1}$, as the storage level at time t, e_t, plus the net energy charged into the battery and minus the energy that is used to drive the vehicle, which is determined by the random state variable x_t. Note that e_{t+1} is written as a function of e_t, x_t^a and u_t^a. Nevertheless, because x_t^a and u_t^a are functions of e_t, x_t and u_t, the energy stored in the battery at time $t+1, e_{t+1}$, could also be written as functions of these. This, however, would complicate the formulation and for this reason it is omitted here.

An interesting case is when the EV-user desires to drive a trip that cannot be covered by the energy presently stored in the battery. In the model, the user starts the trip disregarding the that there is not enough energy on the battery and is thus forced out into a charging-discharging process until the destination is reached. Every time the battery is drained in the process, an unavailability event occurs and the penalty is paid. What is important here, though, is that this charging-discharging process yields the same aggregate penalty as if the user were to charge the vehicle to the required level before embarking on the trip.

3.5. Constraints

The desired charging of the battery is limited to being within the bounds for the rate of charge:

$$u_{\min} \leqslant u_t \leqslant u_{\max}. \quad (11)$$

The storage level on the battery is similarly constrained to being within the storage limits of the battery:

$$e_{\min} \leqslant e_t \leqslant e_{\max}. \quad (12)$$

3.6. Objective function

The revenue at time period t is given by:

$$R_t(S_t, u_t) = -\lambda_t \omega u_t^a - \mathbb{1}_{\{x_t \in S_D, e_t = e_{\min}\}} \omega \phi. \quad (13)$$

The first term, $\lambda_t \omega u_t^a$, is the cost incurred from charging the vehicle. The second term, $\mathbb{1}_{\{x_t \in S_D, e_t = e_{\min}\}} \omega \phi$, is the penalty incurred when the user desires to use the vehicle, but he/she cannot do so, because there is not enough energy stored in the battery. Note that this happens precisely when $x_t \neq x_t^a$. Note also that the revenue is equal to the sum of the costs and the penalty with a negative sign.

The revenue at the end of the optimization horizon, i.e. at time T, is given by:

$$R_T(S_T) = \eta_d e_T \frac{1}{T}\sum_{t=1}^{T} \lambda_t. \quad (14)$$

This equation sets the terminal revenue as the profit that could be made by selling the remaining energy in the battery at the average observed price. One could argue for the use of other terminal conditions: For example, we could replace $\frac{1}{T}\sum_{t=1}^{T}\lambda_t$ in (14) with either λ_T or $\max_t\{\lambda_t\}$. However, we find it more appropriate to use $\frac{1}{T}\sum_{t=1}^{T}\lambda_t$, as this reflects the average economic value of the energy remaining in the battery at $t = T$ if history repeats itself. Besides, $\frac{1}{T}\sum_{t=1}^{T}\lambda_t$ constitutes a better prediction of the future electricity price than λ_T, and $\max_t\{\lambda_t\}$ would probably lead to an over-estimation of the economic value of the leftover energy, since the battery cannot be fully discharged instantly, even if the maximum electricity price encourages the EV user to do so. A terminal condition is important in obtaining a solution for this problem, as this condition provides the starting point from which the resulting stochastic dynamic programming model is solved using backward induction [23]. However, as explained later, the proposed algorithm is to be applied within a rolling-horizon decision-making process, and as a result, the impact of the terminal condition on the charging pattern is conveniently lessened.

Let $\mathcal{U}_s$ denote the set of feasible decisions according to Eqs. (8)–(12), when the system is in state s. Let Π denote the set of all feasible policies. A policy, π, is a collection of decisions $u_t^\pi(s) \in \mathcal{U}_s$, spanning the horizon from $t = 0$ to $t = T$ and all states S_t. Thus for each t and each state $S_t, \pi \in \Pi$ will contain the action, $u_t^\pi(S_t)$, under the policy π. For each $\pi \in \Pi$, we can now define the total expected revenue of that policy from time t to T as:

$$J_t^\pi(S_t) = \mathbb{E}\left[\sum_{\tau=t}^{T} R_\tau(S_\tau, u_\tau^\pi(S_\tau))\middle| S_t\right], \quad (15)$$

where T is the optimization horizon. The objective is then to find a policy, π^*, that satisfies:

$$J_t^{\pi^*}(S_t) = \sup_{\pi \in \Pi} J_t^\pi(S_t), \quad (16)$$

for all $0 \leqslant t \leqslant T$.

3.7. Solution algorithm

Finding an exact solution to the problem stated in (16) will be difficult in general due to the randomness and the continuous nature of states and decisions. As the decision at time t depends on the decisions and the values of random variables in previous time periods, the problem grows exponentially as the number of time steps is increased. In order to capture the actual driving patterns and to integrate them into the model in a sensible manner, it is essential that the time resolution is high (1-min or 5-min intervals). Due to the fluctuating electricity price and the diurnal variation in the driving pattern, the horizon should be a minimum of one-day ahead. If this is to be accomplished, an exponential growth of the problem is not viable.

Instead we solve the problem by discretizing states and decisions. This yields a discrete stochastic dynamic programming problem that is solved using backward induction and Bellman's principle of optimality. As the driving states are already discrete, the level of energy in the battery and the decision variable u_t remain to be discretized. Suppose that the energy stored in the battery e_t is discretized into M states and there are N driving states. This yields a total of $N \times M$ possible state values for each time step. We define I_t as the index set of possible values that the discretized state variable, $\widetilde{S}_t$, can take on. We now define the Bellman equation for the problem stated in Eq. (16) as

$$V_t(\widetilde{S}_t) = \max_{\tilde{u}_t \in \widetilde{U}(\widetilde{S}_t)} \left\{R_t(\widetilde{S}_t, \tilde{u}_t) + \beta \mathbb{E}_t\left[V_{t+1}(\widetilde{S}_{t+1})|\widetilde{S}_t\right]\right\} \quad (17)$$

$$= \max_{\tilde{u}_t \in \widetilde{U}(\widetilde{S}_t)} \left\{R_t(\widetilde{S}_t, \tilde{u}_t) + \beta \sum_{i \in I_{t+1}} \mathbb{P}_t(\widetilde{S}_{t+1}^i|\widetilde{S}_t) V_{t+1}(\widetilde{S}_{t+1}^i)\right\}, \quad (18)$$

where $\widetilde{S}_t$ is the set of discretized values of S_t and $\widetilde{U}(\widetilde{S}_t)$ is the set of discretized possible actions in state $\widetilde{S}_t$. As the exogenous random variable X_t is defined by a Markov chain, the Bellman equation in Eq. (18) represents a Markov decision process, which can be solved using backwards induction, as sketched in Table 1.

Table 1
Pseudocode using backwards induction to obtain the optimal policy $\tilde{\pi}^*$ as the collection of $\tilde{u}^*_{s,t}$.

Backwards Induction Pseudocode
Initialize: The terminal value is defined as $V_T(\tilde{S}_T) = R_T(\tilde{S}_T)$ given by (14)
for $t = T-1$ to 0 **do**:
for $s \in I_t$ **do**:
$\tilde{u}^*_{s,t} = \arg\max_{\tilde{u}_t \in \tilde{U}(s)} \left\{ R_t(s,\tilde{u}_t) + \beta \sum_{i \in I_{t+1}} \mathbb{P}_t(i\|s) V_{t+1}(i) \right\}$
$V_t(s) = R_t(s,\tilde{u}^*_{s,t}) + \beta \sum_{i \in I_{t+1}} \mathbb{P}_t(i\|s) V_{t+1}(i)$
end s **for**
end t **for**

Using the algorithm in Table 1, we find the optimal discretized policy $\tilde{\pi}^*$ as the collection of $\tilde{u}^*_{s,t}$ indicating when and how much to charge, depending on both the time t and the pair s of driving and battery states.

It is advisable to run the algorithm over a long horizon, say two days, to incorporate the diurnal variation in the driving pattern and in the energy price. In addition, the longer the horizon covered by the optimization process, the smaller the influence of the terminal condition. Indeed, we propose to re-run the algorithm in Table 1 for every time step, with a horizon that is extended accordingly, following a rolling-window process.

4. Results and discussion

In this section the model has been implemented and run for an electric vehicle with characteristics similar to those of a Nissan Leaf. The data at our disposal only includes two states, *driving* and *not driving*. We consider a Markov model with three states; one time-varying *not driving* state and two time-invariant *driving* states. Model results are compared with those obtained from "rule-of-thumb" policies to assess the economic performance of the proposed decision-support tool. We first present an in-sample study with the model fitted to the training set, which serves to illustrate its main features. We then carry out an out-of-sample study to evaluate the performance of the model on the test set. For simplicity, we assume that the vehicle is plugged into the electricity grid when not driving.

4.1. Model characteristics

We consider an EV with a battery capacity $\kappa = 24$ kW h and an average consumption of $\mu_i = 0.20$ kW h/km. The entire battery capacity is assumed to be available for use, i.e. $e_{\max} = 24$ kW h and $e_{\min} = 0$ kW h. We also consider that the vehicle is mainly to be employed in an urban driving cycle with an average speed of $v_i = 40$ km/h, including stopping for red lights and congestion. This yields a range of 120 km on one charge, a drive time of 3 h, and an average power consumption of 8 kW. Regarding the charging, we assume a maximum charging capacity of $u_{\max} = 4$ kW (typical maximal power infeed to a residential household). In the base case the vehicle is not allowed to discharge power back into the grid, i.e. $u_{\min} = 0$ kW. This case is subsequently extended to allow for discharging via a V2G scheme with $u_{\min} = -4$ kW. The charging efficiency parameters are $\eta^c = \eta^d = 0.9$. On the basis of these characteristics, the vehicle resembles the Nissan Leaf, which is one of the top selling EVs in the world (as of January 2013).

We consider an optimization horizon covering 48 h in advance to incorporate the diurnal variation in the energy price as well as in the driving pattern. Furthermore, as already explained in Section 3.7, the relatively long horizon is used to decrease the influence of the terminal condition on the optimal charging scheme, which is gradually obtained from the rolling-window process. The time resolution in the model is in minutes, which yields a total of 2880 time steps. A 1-min time resolution is chosen to adequately model the use of the vehicle. As we consider a horizon of 48 h, the discount factor is set to $\beta = 1$. The state variable for the energy charged on the battery is discretized into 360 different states. Likewise, the state variable for the use of the vehicle has three states (one *not driving* and two *driving* states). Therefore, the model relies on 3×360 different states for each time step. In the base case, where only charging is allowed, the vehicle charges at either full rate or not at all, thus the decision variable u_t can only take on two different values. The optimal solution is found in less than a minute on a personal computer with a 2.70 GHz processor and 8.0 GB RAM, which is satisfactory. The model can be straightforwardly modified to work with 5-min or 10-min time resolution with a view to further decreasing the solution time. Also, the discretization of the energy charged on the battery can be coarser. This may be useful if the model is extended with more driving states, or the model has to be implemented with less computing capacity, or if the optimization horizon has to be extended. However, as the model run-time is quite small, such efforts have not been pursued. We notice that the model is parameterized in terms of the penalty in €/h, ϕ, incurred when the vehicle does not have enough energy in the battery to complete the desired trip. This can also be seen as a risk-aversion parameter, where the risk of not completing a trip is weighed against minimizing the costs of driving.

With regard to the electricity price, we use the Nordpool DK1 spot-price historical series. We consider that the EV charging controller receives a 48-h price forecast from, for example, a distribution system operator (DSO). The Nordpool spot price is determined each day in blocks of 24 hourly values and is made public at noon the previous day. Therefore we assume that the risk associated with the volatility of the electricity price is handled by the DSO or some other intermediary, but not by the end-consumer. Besides, the 48-h price forecast may be updated, if appropriate, every time (every minute) the model is re-run as part of the rolling-window process.

4.2. In-sample study

Next we use the training data set defined in Section 2.3 to estimate the transition probability matrix of the inhomogeneous hidden Markov model that we use to describe the driving patterns, as explained in Sections 2.1 and 2.2. Then we simulate plausible driving scenarios based on this model and evaluate the performance of the proposed decision-support tool for optimal charging on these scenarios. Therefore, the analysis carried out here is in-sample, i.e. it assumes that the fitted stochastic model for driving patterns perfectly captures the actual nature of the use of the vehicle. The purpose of this study is then to illustrate the main features of the proposed decision-making tool. First, we analyze schemes where only charging is permitted. Then we consider V2G schemes [12,18], where the vehicle is permitted to supply power from the battery to the grid. We use electricity prices from 00:00 on the 25-01-2012 to 00:00 on the 29-01-2012.

4.2.1. Charging-only schemes

Fig. 4 shows the estimated time-varying probability of starting a trip, the electricity price, and selected values for the optimal policy $\tilde{\pi}^*$, which defines the appropriate charging action to be undertaken given the state $\tilde{S}_t$ and the time t, i.e., the optimal policy indicates whether the EV should be charged or not at time t given the energy level of its battery at that time t. The optimal policy may take values in the set $\{1, 0\}$ for charging and not charging, respectively. In Fig. 4 the battery state is indicated on the vertical axis for different levels of charge, expressed as a percentage of the battery capacity

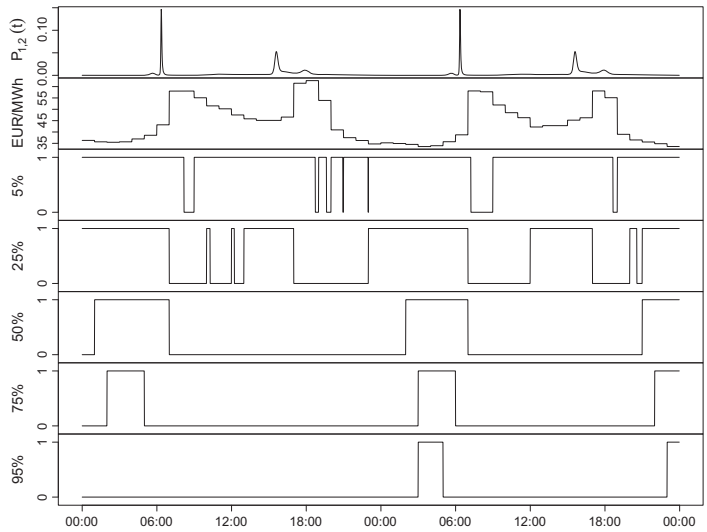

Fig. 4. From top to bottom: time-varying probabilities of starting a trip, electricity price, and $\tilde{\pi}^*$ for different levels of charge on the battery and penalty $\phi = 10$ €/h.

e_{max}. The time is indicated on the horizontal axis. Additionally, note that Fig. 4 only shows the charging decisions for when the vehicle is not driving, as we assume that it is not possible to stop a trip and recharge, unless the battery is fully depleted. It is important to stress that Fig. 4 shows a single run of the optimization algorithm. The difference between this snap-shot of the algorithm and the rolling window process will be illustrated subsequently.

It can be observed from this figure that if the energy level of the battery is 5%, the optimal decision is to always charge, except in those time periods when the probability of driving is low and the electricity price is particularly high. In contrast, if the energy level of the battery is 50%, the vehicle is only charged when the energy price is comparatively low. This charging policy becomes more extreme as the level of charge approaches 100%. Indeed, if the energy level of the battery is equal to 95% of e_{max}, the EV is only charged in those time periods where the energy price is expected to reach its lowest values.

In reality, the proposed charging algorithm is to be used following a rolling-horizon process, which allows for updating the energy price forecast and reducing the effect of the terminal condition on the optimal policy, as highlighted next. In Fig. 5 the results yielded by the algorithm when implemented over a fixed two-day horizon are compared to those obtained considering a two-day rolling-horizon. In the rolling-horizon optimization, the model is rerun every hour and the optimal policy updated accordingly. The rolling-horizon is kept fixed to two days in advance, and consequently we use energy prices from 00:00 on the 25-01-2012 to 00:00 on the 29-01-2012, that is, four days in total. From Fig. 5 we see that there are only slight deviations between the rolling horizon and the fixed horizon procedures within the first day. On the second day, however, we begin to see deviations that go beyond a single spike. As the time approaches time T, more discrepancies are observed between the two models, indicating that the terminal condition has an impact on the optimal charging policy. However, this impact is mostly confined to the last time periods of the 48-h optimization horizon, and therefore implementation of the proposed algorithm in a rolling-horizon fashion can reduce, if not completely eliminate, this effect.

To illustrate the actual implementation of an optimal policy $\tilde{\pi}^*$, we run the following simulation process. First, $\tilde{\pi}^*$ is computed and defined for different values of the penalty ϕ. Second, a plausible realization of the driving pattern is simulated and the corresponding time evolution of the level of charge on the battery is determined according to this realization and the optimal policy $\tilde{\pi}^*$. The results of such a simulation are shown in Fig. 6. It can be seen that as the penalty increases, the level of charge on the battery is correspondingly higher, conditional on the same realization of the driving pattern.

Another promising aspect of EVs is the possibility of supplying power into the grid at times of high demand. This is investigated in the following section.

4.2.2. Vehicle-to-grid schemes

Allowing for the vehicle to supply power from the battery into the grid has the potential to help mitigate the effects of peak power demand. This operation mode is usually referred to as a Vehicle-to-Grid (V2G) scheme. A V2G scheme is investigated here from the perspective of a single vehicle.

Implementation of the V2G scheme is by setting u_{min} to -4 kW and keeping all other parameter values unchanged. The optimal policy, $\tilde{\pi}^*_{V2G}$, obtained by implementing a V2G scheme for a penalty value of $\phi = 10$ €/h, is shown in Fig. 7, which is similar to Fig. 5, except that the optimal policy may take values in the set $\{1, 0, -1\}$ for charging, not charging, and discharging, respectively. Observe that when the energy level in the battery, e_t, is low, the optimal policy, $\tilde{\pi}^*_{V2G}$, basically consists in charging at almost every time t, except in those periods with electricity prices at their peak and a low probability of driving. As the energy level in the battery increases, the policy changes to supplying power into the grid at the price peaks and to charging at the price valleys. The proposed

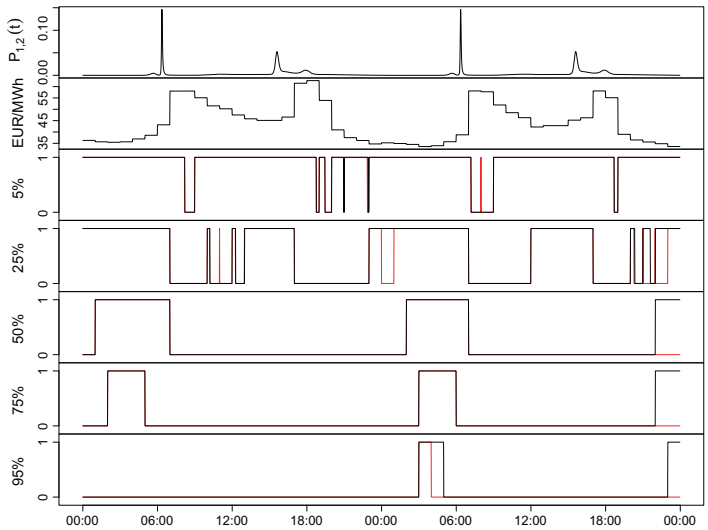

Fig. 5. Analogous to Fig. 4. A rolling-horizon policy has been implemented and the resulting policy is shown in red along with the fixed-window policy in black in the bottom five graphs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

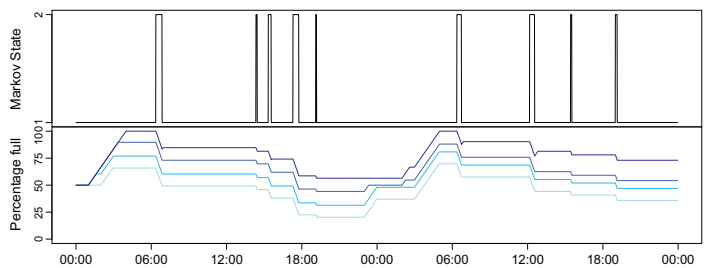

Fig. 6. Top: realization of the driving pattern. Bottom: the corresponding charge on the battery in percent for different penalty values when implementing π^*. The lightest blue line refers to the lowest penalty and the penalty increases with the darker shades of blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

V2G algorithm thus weighs the costs associated with running out of charge on the battery against the gains from delaying charging to when the energy price is low and the gains from supplying power into the grid when the energy price is high. Also in Fig. 7, notice that the optimal policy shows some "spikes" where the optimal decision is changed from charging to not charging for a short time. This is linked to the fact that the electricity price traded on Nordpool is in hourly time resolution, and therefore the price changes only every hour and the corresponding price change may be large. As the vehicle decides the appropriate action for every minute, it is able to exploit this in its charging strategy.

4.3. Out-of-sample study

We now evaluate the model performance on the test data set defined in Section 2.2. Therefore, we provide results from testing the optimal charging policies on the actual utilization of the vehicle in the second half of the data period. This study is thus performed out of sample. We use actual electricity prices that were observed in the Danish area DK1 of the Noordpool market during the time period spanning from 00:00 on 01-01-2011 to 00:00 on 08-03-2011. Consequently, this study seeks to estimate how much it would have cost to run an electric vehicle in DK1 during such a time period, given the driving patterns analyzed in Section 2 and on the assumption that the electric vehicle is directly exposed to the DK1-market prices (e.g., through the DSO, a retailer or an aggregator).

Fig. 8 shows the state of charge of the battery for different charging policies. Note that this figure is analogous to the bottom plot of Fig. 6, except that different charging policies are considered and the time shown is 65 days. In particular, the top graph in Fig. 8 is obtained using the proposed decision-support tool to find the

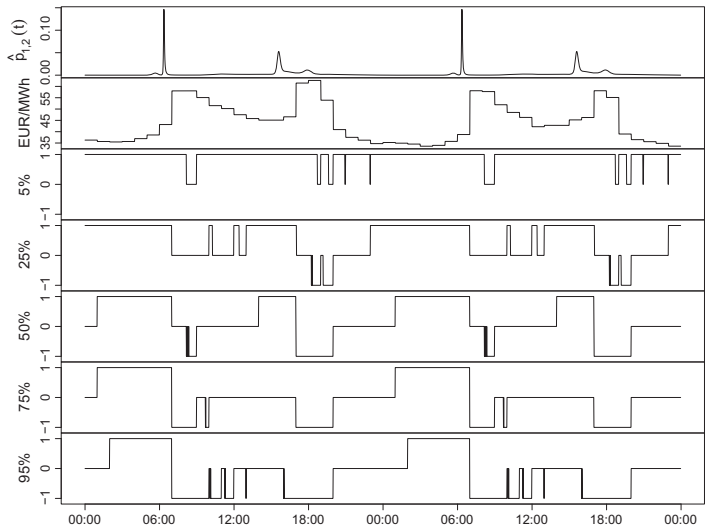

Fig. 7. Top to bottom: time-varying probabilities of starting a trip, electricity price, and $\hat{\pi}^{+}_{V2G}$ for different levels of charge on the battery with penalty $\phi = 10$ €/h. In this case, the discharge of the EV battery back into the grid is considered and therefore, the policy can take on negative values. This figure is similar to Fig. 5, where only charging is allowed.

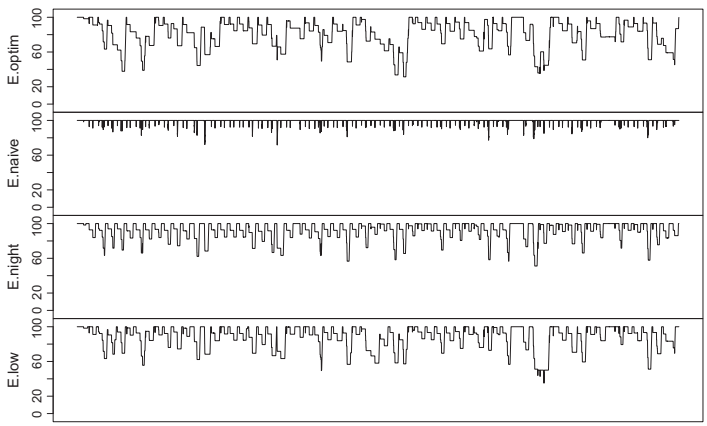

Fig. 8. State of charge in % for different charging policies, from top to bottom: Optimized charging with $\phi = 10$ €/h (with only charging allowed), naive charging, night charging and low-price charging.

optimal policy for charging (V2G operation mode is not permitted). The lower three graphs represent different rule-of-thumb policies. We refer to the first one as "naive charging", according to which the vehicle is charged immediately upon being parked. The second to last is called "night charging", and entails charging the vehicle at night between 10 pm and 6 am or if the charge on the battery is below 50%. The last one is "low price charging", under which the vehicle is charged if the electricity price is in the lowest 20%-quantile of the price distribution for the next 24 h or if the charge on the battery is below 50%. From Fig. 8, notice that none of the strategies empties the battery at any time.

Fig. 9 is similar to Fig. 8, except that in this case the V2G operation mode is allowed. The optimal charging policy is compared to two other rule-of-thumb policies in which V2G is permitted. We refer to these two rule-of-thumb policies as "unbounded" and "bounded" V2G schemes. In the unbounded V2G scheme, the

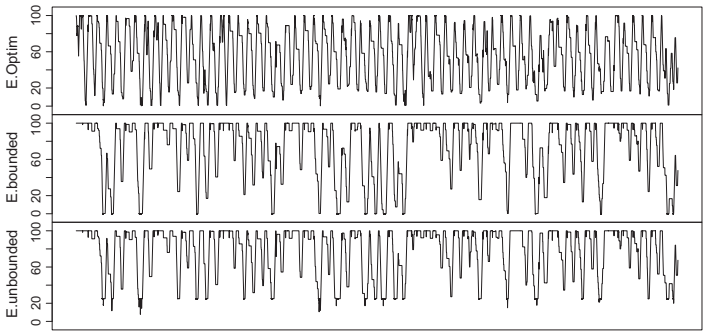

Fig. 9. State of charge in % for different charging policies, from top to bottom: Optimized charging with $\phi = 10$ €/h (with V2G charging allowed), rule-of-thumb V2G unbounded, rule-of-thumb V2G bounded.

vehicle is charged if the electricity price is in the lowest 30%-quantile of the price distribution for the next 24 h and discharged if the electricity price is above the 90%-quantile. The bounded V2G scheme works the same as the unbounded one, except for the fact that in the former the vehicle is also charged if the energy level of the battery goes below the 25% of its maximum capacity.

Comparing Figs. 8 and 9, it is seen that the total battery capacity is exploited when using a V2G scheme. This has the side effect of the battery being depleted at some times during the day. Consequently, the vehicle is not able to cover the user's driving needs, should the user desire to drive. This can, however, easily be mitigated by setting an artificial limit on how much the battery can be depleted. In practice, this is done by setting e_{min} in Eq. (12) higher than the actual minimum storage level of the battery.

Table 2 compares the costs and availability of the vehicle under the different charging policies. We consider the average daily cost of running the vehicle in € and the number of events where the vehicle is not able to cover the user driving needs counted over the 65 days that the test data set spans.

Let us consider first the strategies in Table 2 under which only charging is allowed, we see that there are no observed events of not having enough charge on the battery to complete a trip. Also, we notice, as expected, that the optimal charging strategies have lower costs than the rule-of-thumb policies. The low-price charging strategy is indeed the rule-of-thumb policy that approximates closest to the optimal policy in terms of costs and availability. It yields, however, an average daily cost which is around 12–24% higher than that obtained from implementing the proposed decision-support tool.

Table 2
Average daily costs in € and number of events where there is enough charge on battery to service user driving needs.

Penalty ϕ (€/h)	Charging only		V2G permitted	
	Cost (€)	Events	Cost (€)	Events
2	0.170	0	−0.097	12
5	0.174	0	−0.084	8
10	0.177	0	−0.061	3
100	0.181	0	−0.047	0
1000	0.188	0	−0.019	0
Naive	0.323	0	–	–
Night	0.284	0	–	–
Low price	0.210	0	–	–
V2G unbounded	–	–	0.071	11
V2G bounded	–	–	0.133	0

When a charging policy allows for V2G operation mode, caution should be exercised to prevent the vehicle from being fully discharged when the end-user wants to drive. Notice, however, that as the penalty is increased, the number of events where the user is not able to drive drops to zero. Introducing a V2G charging scheme allows for substantially reducing the cost associated with driving as opposed to charging-only schemes, and may even result in negative average costs. Observe that the optimal charging policy developed in this paper clearly outperforms the rule-of-thumb V2G schemes. In the unbounded case, charging costs are substantially reduced, but multiple out-of-battery events are recorded. Imposing a lower bound on the discharging solves this problem, but at the expense of considerably increasing the running cost of the vehicle, to such an extent that it nearly doubles.

The difference in performance between the optimal charging strategy and the rule-of-thumb policies can be expected to become larger for electric vehicles covering higher distances or with lower battery capacity.

Lastly, we would like conclude this section by pointing out that, in general, the spot price is not the price observed by the end-user. Indeed, the end-user faces a price that includes taxes and other costs on top of the spot electricity price. As an example, consider a country like Denmark, where the average price of electricity paid by the end-user, including taxes and fees, is around €300/MW h, which is 5–10 times the average spot price [24]. In the current Danish power system, fees and taxes are imposed on the amount of electricity consumed by the end-user, not on its total cost. This does not encourage the end-user to switch to a smart consumption of energy based on variable prices. In fact, if the taxes and fees were implemented as a function of the total energy cost, the savings from switching to a smart charging policy in Denmark could be multiplied by a factor of between five and ten.

5. Conclusion

This paper proposes an algorithm to optimally charge an electric vehicle that accounts for the uncertainty in the user's driving patterns. The algorithm is built on an inhomogeneous (hidden) Markov chain model that provides the probability of the vehicle being in use at any time of the day and captures the varying trip durations. Stochastic dynamic programming is then used to determine the optimal charging policy depending on the use of the vehicle, the risk aversion of the end-user, and the electricity price.

The proposed charging model is fitted to a training dataset spanning approximately two and a half months. An in-sample

study is first carried out to: (i) investigate the impact of the termination condition on the resulting charging policy, (ii) highlight the advantages of a rolling-horizon implementation of the algorithm, (iii) assess the effect of an unavailability cost on the model output, and (iv) illustrate the operation of the EV in vehicle-to-grid mode.

The performance of the model is then evaluated out of sample on a test dataset that spans the two succeeding months. and includes the electricity prices from the Danish area DK1 of Nordpool. The out-of-sample study shows that the costs associated with running the vehicle are decreased significantly when the charging strategy is determined by the proposed optimization model, with little or no inconvenience to the end-user. More specifically, the daily cost savings range from approx. 19–47% with respect to a variety of rule-of-thumb charging strategies. Running costs can be reduced even further if the vehicle is permitted to supply power into the grid. Indeed, our numerical results indicate that running costs can be turned into net profit under an optimized V2G scheme, with savings that amount up to 135% with regard to a naive utilization of the V2G functionality for the case in which no unavailability events are allowed.

The proposed stochastic dynamic programming model for EV charging is versatile and can easily be adapted to any specific vehicle, thus providing a customized charging policy that enables the EV-user to save on running costs or even to make profit under a vehicle-to-grid scheme.

A possible extension would be to apply the proposed model to data with more Markov states, which could be used to investigate the benefits of installing more public charging stations as opposed to home charging, or to capture different driving states such as "urban", "rural", or "highway". With a view to its practical application, the proposed Markov decision model could be upgraded to handle transition probabilities that are estimated adaptively in time. Indeed, adaptivity is key to capturing structural changes in the driving behavior of the EV user, for example, those that could follow from the EV user buying another vehicle or moving to a new place.

Further research could be also directed at modeling a fleet of vehicles by using a mixed-effects model. The optimization scheme could be applied individually to each vehicle and the total population load could be evaluated. This would highlight if and how EVs could be used to mitigate an increase in peak electricity demand when switching from combustion-based vehicles to EVs. Other investigations could focus on the relationship between EVs and renewable energy sources and how EVs could be used to move the excess production to time periods of high demand, possibly making renewables more economically competitive. Finally, it would also be interesting to assess how the cost estimates provided in the out-of-sample study in Section 4 are extrapolated to other electricity market with a different price profile (e.g., one with a much less seasonal component due to a high penetration of solar PV capacity).

Acknowledgments

DSF (Det Strategiske Forskningsråd) is to be acknowledged for partly funding the work of Emil B. Iversen, Juan M. Morales and Henrik Madsen through the Ensymora Project (No. 10-093904/DSF). Furthermore, Juan M. Morales and Henrik Madsen are partly funded by the iPower platform project, supported by DSF (Det Strategiske Forskningsråd) and RTI (Rådet for Teknologi og Innovation), which are hereby acknowledged. Finally, we thank DTU Transport for providing the data used in this research.

References

[1] Green R, Wang L, Alam M. The impact of plug-in hybrid electric vehicles on distribution networks: a review and outlook. In: IEEE PES general meeting; 2010. p. 1–8.
[2] Bessa RJ, Matos MA. Economic and technical management of an aggregation agent for electric vehicles: a literature survey. Eur Trans Electr Power 2012;22:334–50.
[3] Weiller C. Plug-in hybrid electric vehicle impacts on hourly electricity demand in the United States. Energy Policy 2011;39:3766–78.
[4] Shao S, Pipattanasomporn M, Rahman S. Grid integration of electric vehicles and demand response with customer choice. IEEE Trans Smart Grid 2012;3:543–50.
[5] Sioshansi R. Modeling the impacts of electricity tariffs on plug-in hybrid electric vehicle charging, costs, and emissions. Oper Res 2012;60:506–16.
[6] Juul N, Meibom P. Optimal configuration of future energy systems including road transport and vehicle-to-grid capabilities. In: EWEC scientific proceedings; 2009. p. 168–74.
[7] Shortt A, O'Malley M. Impact of optimal charging of electric vehicles on future generation portfolios. In: IEEE PES/IAS conference on Sustainable Alternative Energy (SAE). IEEE; 2009. p. 1–6.
[8] Kiviluoma J, Meibom P. Influence of wind power, plug-in electric vehicles, and heat storages on power system investments. Energy 2010;35:1244–55.
[9] Gan L, Topcu U, Low S. Optimal decentralized protocol for electric vehicle charging. Technical Report. California Institute of Technology; 2011. <http://www.cds.caltech.edu/utopcu/images//9/96/GTL-cdc11.pdf>.
[10] Ma Z, Callaway D, Hiskens I. Decentralized charging control for large populations of plug-in electric vehicles: application of the Nash certainty equivalence principle. In: IEEE international Conference on Control Applications (CCA); 2010. p. 191–95.
[11] Richardson P, Flynn D, Keane A. Optimal charging of electric vehicles in low-voltage distribution systems. IEEE Trans Power Syst 2012;27:268–79.
[12] Kempton W, Tomic J. Vehicle-to-grid power fundamentals: calculating capacity and net revenue. J. Power Sources 2005;144:268–79.
[13] Rotering N, Ilic M. Optimal charge control of plug-in hybrid electric vehicles in deregulated electricity markets. IEEE Trans Power Syst 2011;26:1021–9.
[14] Kristoffersen TK, Capion K, Meibom P. Optimal charging of electric drive vehicles in a market environment. Appl Energy 2011;88:1940–8.
[15] Momber I, Gomez T, Soder L. PEV fleet scheduling with electricity market and grid signals. In: 10th international conference on the European Energy Market (EEM); 2013. p. 1–8.
[16] Lojowska A, Kurowicka D, Papaefthymiou G, van der Sluis L. From transportation patterns to power demand: Stochastic modeling of uncontrolled domestic charging of electric vehicles. In: IEEE PES general meeting; 2011. p. 1–7.
[17] Iversen EB, Møller JK, Morales JM, Madsen H. Inhomogeneous Markov models for describing driving patterns, technical report. Technical University of Denmark; 2013. <http://orbit.dtu.dk/ws/files/51822610/4.pdf>.
[18] Kempton W, Letendre SE. Electric vehicles as a new power source for electric utilities. Transport Res Part D–Transport Environ 1997;2:157–75.
[19] Grimmett G, Stirzaker D. Probability and random processes. Texts from Oxford University Press. Oxford University Press; 2001.
[20] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer series in statistics. Springer; 2008.
[21] Madsen H, Thyregod P. Introduction to general and generalized linear models. Chapman & Hall/CRC Texts in Statistical Science. Chapman & Hall; 2010.
[22] Zucchini W, MacDonald I. Hidden Markov models for time series: an introduction using R. Monographs on statistics and applied probability. Taylor & Francis Group; 2009.
[23] Puterman ML. Markov decision processes: discrete stochastic dynamic programming, vol. 414. Wiley; 2009.
[24] European Commission. Europe's energy portal; 2012. <http://www.energy.eu/#Domestic-Elec>.

Paper C

Probabilistic Forecasts of Solar Irradiance using Stochastic Differential Equations

---

Published in *Environmetrics, 2014*.

Environmetrics **Research Article**

Received: 30 August 2013, Revised: 23 January 2014, Accepted: 24 January 2014, Published online in Wiley Online Library: 14 April 2014

(wileyonlinelibrary.com) DOI: 10.1002/10.1002/env.2267

Probabilistic forecasts of solar irradiance using stochastic differential equations

E. B. Iversen, J. M. Morales, J. K. Møller and H. Madsen

Probabilistic forecasts of renewable energy production provide users with valuable information about the uncertainty associated with the expected generation. Current state-of-the-art forecasts for solar irradiance have focused on producing reliable *point* forecasts. The additional information included in probabilistic forecasts may be paramount for decision makers to efficiently make use of this uncertain and variable generation. In this paper, a stochastic differential equation framework for modeling the uncertainty associated with the solar irradiance point forecast is proposed. This modeling approach allows for characterizing both the interdependence structure of prediction errors of short-term solar irradiance and their predictive distribution. Three different stochastic differential equation models are first fitted to a training data set and subsequently evaluated on a one-year test set. The final model proposed is defined on a bounded and time-varying state space with zero probability almost surely of events outside this space. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: forecasting; stochastic differential equations; solar power; probabilistic forecast; predictive distributions

1. INTRODUCTION

The operation of electric energy systems is today challenged by the increasing level of uncertainty in the electricity supply brought in by the larger and larger share of renewables in the generation mix. Decision-making, operational, and planning problems in electricity markets can be characterized by time-varying and asymmetric costs. These asymmetric costs are caused by the need to continuously balance the electricity system to guarantee a reliable and secure supply of power. An understanding of the underlying uncertainty is, therefore, essential to satisfactorily manage the electricity system. This introduces the need for forecasts describing the entire variation of the renewable generation.

Solar irradiance is a source of renewable energy and, along with wind and hydro, is taking shape as a potential driver for a future free of fossil fuels. The worldwide installed capacity of photovoltaic energy systems has seen a rapid increase from 9.5 GW in 2007 to more than 100 GW by the end of 2012 (European Photovoltaic Industry Association (2013)). The energy generation from solar irradiance is subject to weather conditions and, as such, it constitutes a variable and uncertain energy source.

Current state-of-the-art forecasting techniques for solar energy have focused on point forecasts, that is, the most likely or the average outcome. Such point forecasts, however, do not adequately describe the uncertainty of the power production. This is recognized by the abundance of significant works on probabilistic forecasting for wind power (see, for example, Pinson *et al.* (2007) and Zhou *et al.* (2013)).

In the literature, a variety of different approaches have been taken to provide reliable solar power point forecasts, on the basis of classic methods for time series analysis and modeling (Huang *et al.* (2013); Boland (2008); Ji and Chee (2011); Bacher *et al.* (2009); Yang *et al.* (2012)), artificial neural networks (Mihalakakou *et al.* (2000); Chen *et al.* (2011); Bhardwaj *et al.* (2013)), cloud motion forecasts (Perez *et al*. 2010), and nontime series statistical models (Ridley *et al.* (2010); Kaplanis and Kaplani (2010)). A review of some of these approaches is found in Pedro and Coimbra (2012). In Lorenz *et al.* (2009), a forecast method that makes use of numerical weather predictions (NWPs) and exploits a clear sky model that accounts for the orientation and tilt of the photovoltaic panel is developed.

Probabilistic forecasting of solar irradiance is, though, in its infancy. One work in this area is the one by Mathiesen *et al.* (2013), where postprocessing of NWPs is applied to obtain probabilistic forecasts. Previous work on stochastic differential equations (SDEs) and solar irradiance is, to the best of our knowledge, limited to Soubdhan and Emilion (2010), which formulates a very simple SDE model for solar irradiance. As a consequence of its simplicity, the model was largely unsuccessful at forecasting. SDEs are fruitfully used for wind power forecasting in Møller *et al.* (2013) by considering state-dependent diffusions and external input.

The methodological approach adopted in this work (SDEs, likelihood estimation, and extended Kalman filter) has been used in other fields of application, for example, Madsen and Holst (1995) (heat dynamics of buildings), Møller *et al.* (2011) (phytoplankton modeling), and Breinholt *et al.* (2011) (sewer systems). Other methodological approaches could, of course, be applied to the SDE estimation problem. In particular, Monte Carlo-based methods are widely used to this end, both in discrete time (see, for example, Dowd (2007) in the context of ecosystems modeling) and in continuous time (see, for example, Nicolau (2002) and Pedersen *et al.* (2008), the latter dealing with the problem of geolocation of fish).

* *Correspondence to: E. B. Iversen, Technical University of Denmark, Asmussens Alle, Building 322, DK-2800 Lyngby, Denmark. E-mail: jebi@dtu.dk*

Technical University of Denmark, Asmussens Alle, Building 322, DK-2800 Lyngby, Denmark

This paper describes a new approach to solar irradiance forecasting based on SDEs. Modeling with SDEs has multiple benefits, among others are the following:

1. SDE models are able to produce reliable point forecasts as well as probabilistic forecasts.
2. Model extensions are easy to formulate and have an intuitive interpretation. We can start with a simplistic model and extend it to a sufficient degree of complexity.
3. We can model processes that are bounded and assign zero probability to events outside the bounded interval, which is essential for correct probabilistic forecasts of solar irradiance.
4. We leave the discrete-time realm of Gaussian innovations and consider instead the more general class of continuous-time processes with continuous trajectories.
5. SDEs span a large class of stochastic processes with classical time series models as special cases.

The rest of this paper is structured as follows: Section 2 gives a general introduction to the proposed SDE framework and describes an estimation procedure. Section 3 first introduces a simple SDE model that tracks an NWP of solar irradiance. Mechanistic extensions to this simple model are subsequently proposed to capture some intuitive properties of the underlying stochastic process. Section 3 then ends by providing an SDE model that overcomes some shortcomings of the mechanistic one that are identified using classic techniques of time series analysis. The three different models are then compared with simple as well as complex benchmarks in Section 4, where, in addition, the performance of the finished model is assessed. Lastly, Section 5 concludes the paper.

2. STOCHASTIC DIFFERENTIAL EQUATIONS

An SDE is a differential equation with one or more stochastic terms resulting in a solution that is in itself a stochastic process. SDEs are used to describe various phenomena driven by a large random component and are especially prominent in mathematical finance (Björk (2009); Mikosch (1998)) and physics (Van Kampen (1992); Adomian (1988)). We give here a very short introduction to SDEs and refer the interested reader to Øksendal (2010) for a thorough and mathematically rigorous discussion on the topic.

Suppose that we have the continuous time process $X_t \in \mathcal{X} \subset \mathbb{R}^n$. In general, it is only possible to observe continuous time processes in discrete time. We observe the process X_t through an observation equation at discrete times. Denote the observation at time t_k by $Y_k \in \mathcal{Y} \subset \mathbb{R}^l$ for $k \in \{0, \ldots, N\}$. Let the observation equation be given by

$$Y_k = h\left(X_{t_k}, t_k, e_k\right), \tag{1}$$

where the variable t_k allows for some form of dependence on an external input at time t_k, $e_k \in \mathbb{R}^l$ that is the random observation error, and $h(\cdot) \in \mathbb{R}^l$ is the function that links the process state to the observation. The simplest form of an observation equation is $h(\cdot) = X_{t_k} + e_k$.

2.1. Definition of stochastic differential equations

In the ordinary differential equation setting, the evolution in time of the state variable X_t is given by the deterministic system equation

$$\frac{dX_t}{dt} = f(X_t, t), \tag{2}$$

where $t \in \mathbb{R}$ and $f(\cdot) \in \mathbb{R}^n$. Complex systems such as weather systems are subject to random perturbations of the input or processes that are not specified in the model description. This suggests introducing a stochastic component in the state evolution to capture such perturbations. This can be carried out by formulating the state evolution as an SDE, as carried out in Øksendal (2010). Thus, we can formulate the time evolution of the state of the process in the form

$$\frac{dX_t}{dt} = f\left(X_t, t\right) + g\left(X_t, t\right) W_t, \tag{3}$$

where $W_t \in \mathbb{R}^m$ is an m-dimensional standard Wiener process and $g(\cdot) \in \mathbb{R}^{n \times m}$ is a matrix function (Øksendal, 2010). Multiplying by dt on both sides of (3), we obtain the standard SDE formulation:

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t \tag{4}$$

Notice that we allow for a complex dependence on t, including external input at time t. Although this form is the most common for SDEs, it is not well defined, as the derivative of W_t, $\frac{dW_t}{dt}$, does not exist. Instead, it should be interpreted as an informal way of writing the integral equation:

$$X_t = X_0 + \int_0^t f\left(X_s, s\right) ds + \int_0^t g\left(X_s, s,\right) dW_s \tag{5}$$

In Equation (5), the behavior of the continuous time stochastic process X_t is expressed as the sum of an initial stochastic variable, an ordinary Lebesgue integral, and an Itō integral.

In a deterministic ordinary differential equation setting, the solution would be a single point for each future time t. In the SDE setting, in contrast, the solution is the probability density of X_t for any state, x, and any future time, t. For an Itō process given by the SDE defined in (4) with drift $f(X_t, t)$ and diffusion coefficient $g(X_t, t) = \sqrt{2D(X_t, t)}$, the probability density $j(x, t)$ in the state x at time t of the random variable X_t is given as the solution to the partial differential equation known as the Fokker–Planck equation (Björk, 2009):

$$\frac{\partial}{\partial t} j(x,t) = -\frac{\partial}{\partial x}\left[f(x,t)j(x,t)\right] + \frac{\partial^2}{\partial x^2}\left[D(x,t)j(x,t)\right] \tag{6}$$

Thus, given a specific SDE, we can find the density at any future time by solving a partial differential equation.

Stochastic differential equations are a general class of processes. This is stated by the Lévy-Itō decomposition, which says that under sufficient regularity conditions, all stochastic processes with continuous trajectories can be written as SDEs (Øksendal, 2010). Hence, many of the ordinary discrete-time stochastic processes can be seen as an SDE being sampled at discrete times, and therefore, SDEs are a generalization of generic time series models in discrete time. Besides Øksendal (2010), other useful introductions to SDEs are, for example, Kloeden and Pearson (1977) and Mikosch (1998).

2.2. Parameter estimation

In this section, we outline how to estimate the parameters in an SDE of a specific form, in particular, with a state-independent diffusion term (a thorough description of the estimation method used here can be found in Jazwinski (2007)). Then, in the following section, we show how to transform a process with state-dependent diffusion term into a process with a unit diffusion term, whereby the estimation procedure discussed in this section can be applied.

Consider the model defined by Equations (1) and (4), that is,

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t \tag{7}$$

$$Y_k = h\left(X_{t_k}, t_k, e_k\right), \tag{8}$$

the parameters of which we desire to estimate. The estimation problem, in a nutshell, can be formulated as follows: find a parameter vector, $\hat{\theta} \in \Theta$, that maximizes some objective function of θ. There are several possible choices for such an objective function. A natural option in this framework is to choose an objective function that maximizes the likelihood of seeing the observations given by $\mathcal{Y}_N = \{Y_0, \ldots, Y_N\}$. This leads to selecting the likelihood function as the objective, that is,

$$L(\theta; \mathcal{Y}_N) = p(\mathcal{Y}_N|\theta) = \left(\prod_{k=1}^{N} p\left(Y_k|\mathcal{Y}_{k-1}, \theta\right)\right) p(Y_0|\theta), \tag{9}$$

where $p(\cdot)$ indicates the probability density of the observation Y_k, which is the convolution of X_{t_k} and e_k. Even though this problem could, in principle, be solved using the Fokker–Planck equation, this is only feasible for systems with simple structures, as it involves solving a complex partial differential equation. Consequently, the estimation procedure, which we shall introduce next, relies on the system having a specific form, namely,

$$dX_t = f(X_t, t)dt + g(t)dW_t \tag{10}$$

$$Y_k = h\left(X_{t_k}, t_k\right) + e_k \tag{11}$$

In the system defined by Equations (10) and (11), we assume that $g(\cdot) \in \mathbb{R}^{n\times n}$ does not depend on the state X_t and that the observation noise is an additive Gaussian white noise, that is, $e_k \sim \mathcal{N}(0, S_k(t_k))$, where $S_k(t_k)$ is some covariance matrix, possibly depending on time. It is clear that restricting $g(\cdot)$ to not depend on X_t limits our model framework severely. As we shall see later on in Section 2.3, this can, to a large extent, be remedied by a transformation using Itō-calculus. The restriction of having additive Gaussian measurement noise should be dealt with by appropriate transformations of the observations. For a detailed study of these transformations, we refer the reader to Box and Cox (1964). Outliers and data contaminations also pose a problem, as they may interfere with the normality assumption. For this reason, data preprocessing should be performed in the presence of outliers and contaminated data. The implementation of the extended Kalman filter in the R package CTSM-R, which is used here, includes a simple outlier detection algorithm (Juhl *et al.* (2013)). For a reference on outlier detection and robust statistics, see Rousseeuw and Leroy (2005). Other methods such as particle filters or the ensemble Kalman filter can also be used to estimate the SDE parameters (Arulampalam *et al.* (2002); Ristic *et al.* (2004)). We choose the extended Kalman filter for its simplicity and speed. Furthermore, as we will see in Section 2.3, we will be able to estimate sufficiently complex SDE models using the extended Kalman filter in conjunction with the Lamperti transform. For a more detailed discussion, see, for example, Jazwinski (2007).

As the system defined by Equations (10) and (11) is driven by Wiener noise, which has Gaussian increments, and the observation noise is Gaussian, it is reasonable to assume that the density of $Y_k|\mathcal{Y}_{k-1}$ can be approximated by a Gaussian distribution. Note that the Gaussian distribution is completely characterized by its mean and covariance. This implies that using the extended Kalman filter, which is linear, is appropriate. For a detailed discussion of the extended Kalman filter, see, for example, Chui and Chen (2009).

The one-step predictions for the mean and variance are defined as

$$\widehat{Y}_{k|k-1} = \mathbb{E}\left[Y_k|\mathcal{Y}_{k-1}, \theta\right] \tag{12}$$

$$R_{k|k-1} = \mathbb{V}\left[Y_k|\mathcal{Y}_{k-1}, \theta\right], \tag{13}$$

where $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ denote the expectation and variance, respectively. The innovation is given by

$$\epsilon_k = Y_k - \widehat{Y}_{k|k-1} \tag{14}$$

Using (12)–(14), we can now write the likelihood function as

$$L(\theta; \mathcal{Y}_N) = \left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\epsilon_k^\top R_{k|k-1}^{-1}\epsilon_k\right)}{\sqrt{\det\left(R_{k|k-1}\right)}\left(\sqrt{2\pi}\right)^l} \right) p(Y_0|\theta), \tag{15}$$

where l is the dimension of the sample space and $(\cdot)^\top$ denotes the vector transpose. Here, ϵ_k and $R_{k|k-1}$ can be computed by means of the extended Kalman filter as shown in the remainder of this section. Further conditioning on Y_0 and taking the logarithm yields

$$\log\left(L(\theta; \mathcal{Y}_N | Y_0)\right) = \frac{1}{2}\sum_{k=1}^{N}\left(\log\left(\det\left(R_{k|k-1}\right)\right) + \epsilon_k^\top R_{k|k-1}^{-1}\epsilon_k\right) + \log(2\pi)\frac{Nl}{2} \tag{16}$$

The estimate of θ can be found by solving the optimization problem

$$\hat{\theta} = \arg\max_{\theta\in\Theta}\left(\log\left(L(\theta; \mathcal{Y}_N | Y_0)\right)\right) \tag{17}$$

We use the extended Kalman filter to obtain the necessary one-step predictions mentioned in Equations (12) and (13) by means of the output predictions:

$$\widehat{Y}_{k|k-1} = h\left(\widehat{X}_{k|k-1}, t_k\right) \tag{18}$$

$$R_{k|k-1} = CP_{k|k-1}C^\top + S_k, \tag{19}$$

where C is the first-order expansion of $h(\cdot)$ in (11), that is, $C = \left.\frac{\partial h}{\partial x_t}\right|_{x=\widehat{X}_{k|k-1}, t=t_k}$, and $P_{k|k-1}$ is the one-step prediction of the covariance of the underlying state, $\widehat{X}_{k|k-1}$. The Kalman gain K_k governs how much the one-step prediction of the underlying state, $\widehat{X}_{k|k-1}$, should be adjusted to form the state update, $\widehat{X}_{k|k}$, from the new observation Y_k. It is given by

$$K_k = P_{k|k-1}C^\top R_{k|k-1}^{-1} \tag{20}$$

The state update is then performed as

$$\widehat{X}_{k|k} = \widehat{X}_{k|k-1} + K_k\epsilon_k \tag{21}$$

$$P_{k|k} = P_{k|k-1} - K_k R_{k|k-1} K_k^\top \tag{22}$$

The state prediction equation is given as

$$\frac{d\widehat{X}_{t|k}}{dt} = f\left(\widehat{X}_{t|k}, t\right), \quad t \in [t_k, t_{k+1}[\tag{23}$$

$$\frac{dP_{t|k}}{dt} = AP_{t|k} + P_{t|k}A^\top + g(t)g(t)^\top, \quad t \in [t_k, t_{k+1}[, \tag{24}$$

where $A = \left.\frac{\partial f}{\partial x_t}\right|_{x=\widehat{X}_{k|k-1}, t=t_k}$, that is, the first-order expansion of the drift term $f(\cdot)$ in (10) and $g(\cdot)$ is the diffusion term from (10).

Initial conditions for the extended Kalman filter are $\widehat{X}_{t|t_0} = x_0$ and $P_{t|t_0} = P_0$, which may either be prespecified or estimated along with the parameters. Hence, the state update is a combination of the previous state estimate and the new information obtained from the kth observation, Y_k.

We use the implementation of the extended Kalman filter from the `R`-package CTSM-R described in Juhl *et al.* (2013), which provides both the one-step predictions of future observations and the implementation of the likelihood function (14). The parameters characterizing the SDE system (7) and (8) are then found by solving Equation (17) using the `nlminb` routine for constrained optimization.

2.3. Itō calculus and the Lamperti transform

We will now discuss how an SDE of the form in Equation (7) can be transformed to the form in Equation (10) to allow for the estimation procedure previously introduced. The fundamental tool for the transformation of SDEs is Itō's lemma, as stated in Øksendal (2010). In the succeeding text, we introduce the one-dimensional Itō formula and the Lamperti transform. The multidimensional Itō formula is covered in Øksendal (2010). For a more detailed description of the Lamperti transform and how to apply it to multivariate processes, see Møller and Madsen (2010).

Theorem 1 (The one-dimensional Itō formula) *Let X_t be an Itō process given by*

$$dX_t = f(X_t,t)dt + g(X_t,t)dW_t \tag{25}$$

Let $\psi(x,t) \in C^2([0,\infty)) \times \mathbb{R}$. Then,

$$Z_t = \psi(X_t,t) \tag{26}$$

is again an Itō process, and

$$dZ_t = \frac{\partial \psi}{\partial t}(X_t,t)dt + \frac{\partial \psi}{\partial x}(X_t,t)dX_t + \frac{1}{2}\frac{\partial^2 \psi}{\partial x^2}(X_t,t)(dX_t)^2, \tag{27}$$

where $(dX_t)^2$ is calculated according to the rules

$$dt \cdot dt = dt \cdot dW_t = dW_t \cdot dt = 0, \qquad dW_t \cdot dW_t = dt \tag{28}$$

The Itō formula stated in Theorem 1 can be used to transform the original process to an SDE with unit diffusion by the Lamperti transform.

Theorem 2 (Lamperti transform) *Let X_t be an Itō process defined as in (25) and define*

$$\psi(X_t,t) = \int \frac{1}{g(x,t)}dx\Big|_{x=X_t} \tag{29}$$

If ψ represents a one-to-one mapping from the state space of X_t onto $\mathbb{R}$ for every $t \in [0,\infty)$, then choose $Z_t = \psi(X_t,t)$. Then, Z_t is governed by the SDE

$$dZ_t = \left(\psi_t\left(\psi^{-1}(Z_t,t),t\right) + \frac{f\left(\psi^{-1}(Z_t,t),t\right)}{g\left(\psi^{-1}(Z_t,t),t\right)} - \frac{1}{2}g_x\left(\psi^{-1}(Z_t,t),t\right)\right)dt + dW_t, \tag{30}$$

where $g_x(\cdot)$ and $g_t(\cdot)$ denote the derivatives of $g(\cdot)$ with regard to x and t, respectively, and ψ_t denotes the derivative of ψ with respect to t.

This result is obtained by applying the Itō formula. For proof of Theorem 1, see Øksendal (2010), and for Theorem 2, the reader is referred to Møller and Madsen (2010). The usefulness of these theorems will be illustrated in the following section.

3. SOLAR IRRADIANCE

Next, we build a model for solar irradiance using the SDE framework and theory presented in the previous section. We start with a simple SDE model that tracks an exogenous NWP of the solar irradiance. We subsequently extend this model in two ways: first, by exploiting our mechanistic understanding of solar irradiance, such as the fact that it is a bounded process, and second, by using techniques of time series analysis to identify deficiencies and consequently improve the model to resolve those. Thus, we end up with an SDE model that captures the basic mechanics of the underlying physical process and that is statistically sound.

3.1. Data

The data set at our disposal belongs to a meteorological station located in the western part of Denmark. The data include hourly observations of irradiance on a flat surface together with predictions for irradiance based on an NWP model from the Danish Meteorological Institute. The NWP provides a 48-h forecast of the irradiance, which is updated every 6 h. We use the most recent forecast in the model. The data covers a period of three years from 01/01/09 to 31/12/11. We divide the period into a training and a test set, with the training set covering the first two years and the test set the last year.

3.2. Model 1: tracking the numerical weather prediction

We start by introducing a simple SDE model for solar irradiance that tracks the NWP provided by the Danish Meteorological Institute, that is,

$$dX_t = \theta_x\left(n_t\mu_x - X_t\right)dt + \sigma_x dW_t \tag{31}$$

$$Y_k = X_{t_k} + \epsilon_k \tag{32}$$

Here, X_t is the actual solar irradiance at time t. In this model and the following, we denote the observed solar irradiance at time t_k by Y_k. Parameter n_t is an external input representing the predicted irradiance at time t. In the model, we have parameter μ_x, which allows for a local scaling of the n_t, such that it does not over or undershoot on average. The parameter θ_x determines how rapidly the model reverts to

the predicted level of irradiance. The system noise is controlled by parameter σ_x. The observation error is denoted ϵ_k and is a stochastic variable with distribution $\mathcal{N}(0, \sigma_\epsilon)$.

3.3. Model 2: mechanistic extensions

Some physical characteristics of the solar irradiance are clearly not captured by Model 1. Among these are issues related to its bounded nature: The solar irradiance is never below zero and also never above some time-dependent upper bound determined by the light emitted by the sun and the season. We will remedy this issue and others with the following model:

$$dX_t = \theta_x \left(\frac{n_t + \beta_x}{\gamma m_t + \delta} \mu_x - X_t \right) dt + \sigma_x X_t (1 - X_t) dW_t \tag{33}$$

$$Y_k = \gamma m_{t_k} X_{t_k} + \epsilon_k \tag{34}$$

To capture the cyclical behavior of the solar irradiance, we extend the simple SDE model defined in (31) and (32) by introducing the maximum irradiance in hour t, m_t, as a scaling factor. Here, we compute m_t according to Bird and Hulstrom (1981), where the refraction and absorption in the atmosphere is set to zero.

The aforementioned process is, however, undefined at night, when $m_t = 0$. To overcome this, we can instead think of X_t as a process that describes the state of the atmosphere and how much solar irradiance there would potentially be allowed through. In this context, it clearly makes sense to have X_t defined at night. Thus, we can solve the issue of having $m_t = 0$ by adding a small constant, say $\delta = 0.01$. Given that the n_t is also equal to zero at night, we introduce another parameter β_x (to be estimated) that is added to n_t such that X_t is not forced to tend to zero at night. Furthermore, it should be noticed that the degree of cloud cover is implicitly taken into account through the previous definition of X_t. Thus, the SDE model (33) and (34) can exploit the level of cloud cover at night to predict the level of cloud cover during the next day. This predictive effect depends on the length of the night, and for this reason, it is convenient that the solar irradiance model (33) and (34) runs over night periods as well.

Also, we note that the proportion of solar irradiance that reaches the surface is naturally bounded between zero and one. To impose these natural bounds on the process X_t, we introduce a state-dependent diffusion term $X_t(1 - X_t)$, such that this term decreases to zero as the process approaches the bounds. This implies that the drift term dominates the process near zero and one. Furthermore, because we assume that $\theta_x > 0$ and have that $0 < \frac{n_t + \beta_x}{m_t + \delta} < 1$, the drift term eventually pulls the process away from the bounds.

We have considered the extra terrestrial irradiance, m_t, as the upper limit for the solar irradiance. It should be clear, however, that this level can never be attained, because there will always be some refraction by the atmosphere. Hence, we can possibly scale down the upper limit, m_t, to improve the model. This is carried out by introducing a factor, γ, on the maximum solar irradiance. Another improvement in this direction would be to consider the air mass (Kasten and Young (1989)), as is typically carried out in clear sky models (Ineichen (2006)). The influence of the air mass is linked to the fact that the distance the light travels through the atmosphere is longer when the sun is near the horizon. This is, however, left for future work.

In the estimation procedure, we have assumed that the noise is nonstate-dependent, which is clearly not the case here. Therefore, we need to work with the Lamperti-transformed process. The Lamperti transformation is given by

$$Z_t = \psi(X_t, t) = \int \frac{1}{g(x,t)} dx \Big|_{x = X_t} = \int \frac{1}{\sigma_x x (1 - x)} dx \Big|_{x = X_t} = -\frac{1}{\sigma_x} \log \left(\frac{1 - x}{x} \right) \tag{35}$$

$$X_t = \psi^{-1}(Z_t, t) = \frac{1}{1 + e^{-\sigma_x Z_t}} \tag{36}$$

Noting that $\psi_t(\cdot) = 0$ and $g_x(x, t) = \sigma_x(1 - 2x)$, we can now make use of the Lamperti transform to obtain the process on the transformed Z-space, which becomes

$$dZ_t = \left(\frac{\theta_x \left(\frac{n_t + \beta_x}{\gamma m_t + \delta} \mu_x - \frac{1}{1 + e^{-\sigma_x Z_t}} \right)}{\sigma_x \frac{1}{1 + e^{-\sigma_x Z_t}} \left(1 - \frac{1}{1 + e^{-\sigma_x Z_t}} \right)} - \frac{\sigma_x}{2} \left(1 - 2 \frac{1}{1 + e^{-\sigma_x Z_t}} \right) \right) dt + dW_t \tag{37}$$

$$Y_k = \gamma m_{t_k} \frac{1}{1 + e^{-\sigma_x Z_t}} + \epsilon_k \tag{38}$$

In the sequel, we shall only state the model in the original domain and not in the Lamperti-transformed domain, as they are equivalent in the sense of yielding the same output.

Models (37) and (38) exhibit some important deficiencies. As an example, consider the autocorrelation function of the studentized residuals, which is shown in Figure 1. Observe that there are significant autocorrelation coefficients at the first lags and around the lag 24, which reveals a periodic (daily) component that is not being captured by the SDE models (37) and (38). This and other shortcomings are fixed in the following section.

3.4. Model 3: extensions based on time series analysis

The rate at which the stochastic process X_t should revert to its predicted level may vary over time for a number of reasons. For instance, the NWP may be more accurate at some times than at others, as acknowledged in Hacker and Rife (2007), among others. Furthermore, in our

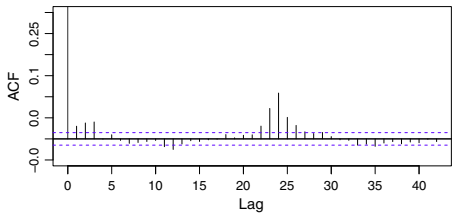

Figure 1. Autocorrelation function for the studentized residuals of Model 2

particular case, the NWP model is only run every 6 h. As it takes around 4 h to solve the NWP model, this leads to the NWP being between 4 and 10 h old, which justifies a varying confidence in it.

To address this issue, we replace the constant θ_x in(33) with the stochastic process A_t, which reverts to the level μ_A. The speed at which this reversion occurs is determined by θ_A, whereas the system noise is governed by σ_A. A_t governs how rapidly X_t tends to its predicted level. This leads to the following model:

$$dX_t = e^{A_t}\left(\frac{n_t+\beta_x}{\gamma m_t+\delta}\left(\mu_x-\omega_1\sin\left(\frac{2\pi}{24}t+\omega_2\right)\right)-X_t\right)dt+\sigma_x X_t(1-X_t)dW_{1,t} \tag{39}$$

$$dA_t = \theta_A(\mu_A-A_t)\,dt+\sigma_A dW_{2,t} \tag{40}$$

$$Y_k = \gamma m_{t_k}X_{t_k}+\epsilon_k, \tag{41}$$

where, in addition, we have introduced a sinusoid to capture the variations over the day in the bias of the NWP. Notice that we work in hourly time steps, which explains $\frac{2\pi}{24}t$. We then introduce a period shift, ω_2, and an amplitude, ω_1. The sinusoid is added to the scaling of the meteorological prediction, which translates into n_t having varying bias over the day. This bias effect is driven by the data available, but it is also acknowledged in Hacker and Rife (2007) among others.

We have ended up with a model that includes a maximum hourly irradiance, an NWP as external input, stochastic time constants, and a non-Gaussian system noise that confines the process between zero and the extra terrestrial irradiance. In the following section, validation results of the final model are presented.

4. MODEL VALIDATION

In this section, the different models are fitted to the data pertaining to the training set and evaluated in terms of their likelihood and information criteria.

The estimation procedure described in Section 2 has been run on a Linux server consisting of 20 processors clocking at 1.8 GHz. A parallelization of optimization problem (17) has been carried out. Results show that it takes approximately 3, 16, and 41 min to estimate Models 1, 2, and 3, respectively, all of them fitted to the training dataset including hourly solar irradiance measurements over two years. For comparison purposes, the most complex benchmark we consider (i.e., the ARX-GLM model we introduce below) is fitted in approximately 8 min. Also, the parallelization has diminishing returns to scale. Indeed, fitting Model 3 takes just over 2 h on a personal computer with a 2.7-GHz processor and 8 GB of RAM.

The estimated parameters for the three SDE models presented in Section 3 are collated in Table 1. As expected from our discussion in that section, μ_x is around 1, β_x is smaller than 0.01, and γ is close to, but smaller than 1. That is, all these parameter estimates are feasible and reasonable considering the physics of the solar irradiance stochastic process, which constitutes a first step in validating the SDE models.

We now compare our SDE models with more classical alternatives. In particular, we consider an autoregressive model with and without external input (referred to as ARX and AR, respectively) and an autoregressive model with external input and time-varying system variability,

Table 1. Parameter estimates for the different models

	$\hat{\theta}_x$	$\hat{\mu}_x$	$\hat{\sigma}_x$	$\hat{\sigma}_\epsilon$	$\hat{\beta}_x$	$\hat{\gamma}$	$\hat{\theta}_A$	$\hat{\mu}_A$	$\hat{\sigma}_A$	$\hat{\omega}_1$	$\hat{\omega}_2$
Model 1	0.699	0.845	113	0.00101	—	—	—	—	—	—	—
Model 2	0.345	0.804	0.701	2.92	0.00409	0.902	—	—	—	—	—
Model 3	—	0.879	0.655	2.89	0.00298	0.887	1.16	−1.08	1.60	0.172	0.116

where the prediction variance is modeled using a generalized linear model (ARX-GLM). Here, we have followed the fitting procedure described in Madsen and Thyregod (2011). The AR model is specified as follows:

$$Y_k = \psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \epsilon_k, \qquad \text{where } \epsilon_k \sim \mathcal{N}\left(0, \sigma^2\right) \tag{42}$$

The ARX model takes the form

$$Y_k = \psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \phi p_k + \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}\left(0, \sigma^2\right) \tag{43}$$

The ARX-GLM model is cast as

$$Y_k = \psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \phi p_k + \epsilon_k, \qquad \tilde{\epsilon}_{k+1} \sim \mathcal{N}\left(0, f_{\tilde{\epsilon}}(\cdot)^2\right) \tag{44}$$

The fitting process reveals that an appropriate form of the variance scaling in the generalized linear model (44) is

$$f_{\tilde{\epsilon}}(k+1) = \sigma \left(m_{t_{k+1}}\right)^{3/4}, \tag{45}$$

where $m_{t_{k+1}}$ is specified as in Model 2. For a general introduction to generalized linear models, see Madsen and Thyregod (2011).

Additionally, we benchmark our SDE models against climatological forecasts. In particular, we consider the following climatological forecasts of the predictive distribution of solar irradiance, listed in order of increasing degree of sophistication:

1. The time-independent empirical distribution of the historical series of solar irradiance.
2. The empirical distribution as a function of "hour-of-day".
3. The empirical distribution as a function of both "hour-of-day" and "month-of-year".

All these three benchmarks are, however, naïve in that they do not exploit the immediately previous observations to predict the distribution of solar irradiance. Besides, as the climatological approach is nonparametric, we use the empirical likelihood to evaluate the fit (Bera and Bilias (2002)). Furthermore, we use the likelihood as the criterion to compare the performance of the different probabilistic forecasting methods. The reason for this is that we are mainly concerned with the conditional distribution of the solar irradiance at a future time. Consequently, traditional point-forecasting metrics, such as the mean absolute error or the root mean square error are not appropriate, because they consider only the deviation from the point forecast. Also, the conditional variance of the solar irradiance changes over time and depending on the state of the process, which is overlooked by both mean absolute error and root mean square error. The likelihood, in contrast, is a proper scoring rule (Gneiting and Raftery, 2007), meaning that a better fit of the data will result in a better score, and this is why we employ the likelihood

Table 2. The log-likelihood of the different models are shown on the training and test set along with information criteria and degrees of freedom

		Training set			Test set
	d.f.	LL	AIC	BIC	LL
Clim.1	—	−96397	—	—	−48421
Clim.2	—	−65060	—	—	−33801
Clim.3	—	−48038	—	—	−24547
AR	6	−50218	—	—	−25172
ARX	7	−49024	—	—	−24567
ARX-GLM	7	−46904	—	—	−23361
Model 1	5	−50286	100582	100621	−25230
Model 2	7	−44413	88840	88894	−22252
Model 3	12	−44162	88348	88441	−22102

The climatological predictors are evaluated in terms of empirical likelihood.
d.f., degrees of freedom; Clim, Climatological benchmark; LL, Log-Likelihood; AIC, Akaike information criterion; BIC, Bayesian information criterion; AR, Autoregressive model; ARX-GLM, Autoregressive model with external input and generalized linear innovations.

both as the objective to be maximized in the fitting procedure and as the performance metric. Moreover, we choose the likelihood over other proper scoring rules such as the continuous rank probability score, as the likelihood in our case is a direct result of the optimization problem stated in Equation (17).

The results from the comparison of the different models are presented in Table 2. In the computation of the likelihood values, we only take into account observations recorded during the day with the ultimate purpose of carrying out a fair comparison. In other words, none of the models in Table 2 are supposed to perform well at nights, when the fitting of a solar irradiance model is meaningless. This holds true for both the benchmarks and the proposed SDE models. We see that Model 3 best describes the data in the training set, as well as in the test set. Furthermore, the improvement from the quite naïve Model 1 to 2 is huge, which justifies the change in the state space (i.e., going from interpreting the state variable X_t as the level of solar irradiance on a surface to interpret it as the proportion of the extra terrestrial irradiance that finds its way through the atmosphere).

We now show the autocorrelation function of the studentized residuals for each of the three SDE models considered. The range of the y-axes is set to $(-0.05, 0.30)$ for a better visual inspection and comparison. In Figure 2, we see that Model 1 has many autocorrelation coefficients that are significant, clearly indicating that this model does not properly capture the dynamics of the solar irradiance process. In contrast, for Model 2, only a couple of the first autocorrelations along with some autocorrelations around lag 24 are still significant. We then end up with Model 3, in which the first 22 autocorrelation coefficients are insignificant in predicting the next time step. However, around lag number 24, we again begin to see significant autocorrelation coefficients. This is most likely caused by local conditions like shadowing (by trees or buildings) or local recurrent weather phenomena such as sea breeze (Bacher *et al.* (2013)).

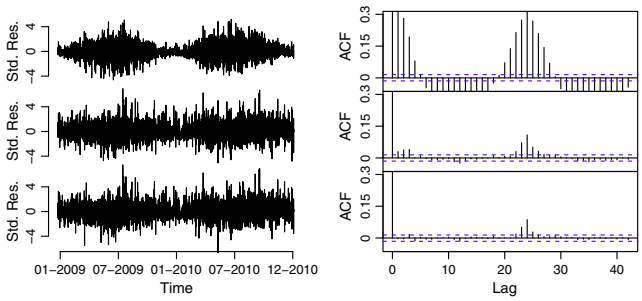

Figure 2. Studentized residuals and autocorrelation function for the residuals. Here, the plots in the top line are from Model 1 continuing to the plots in the bottom line from Model 3

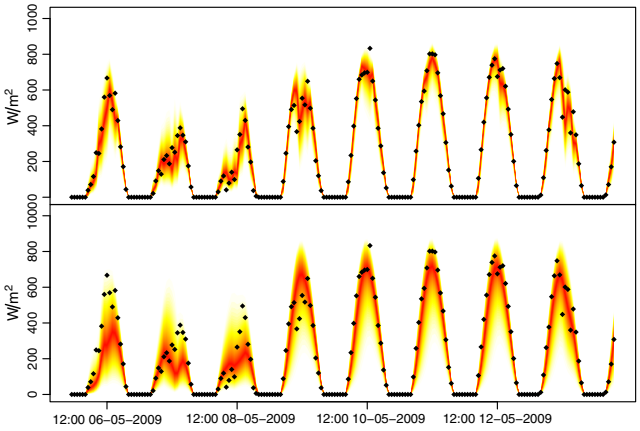

Figure 3. Observations of solar irradiance plotted together with the 1-h ahead (top) and the 24-h ahead (bottom) predictive densities yielded by SDE Model 3. Warmer colors indicate a higher probability of seeing this realization

The output of the SDE models is the conditional predictive density of solar irradiance at each point in time. We obtain these predictive densities by Monte Carlo simulation. However, other numerical procedures are also possible, such as solving Equation (6). In Figure 3, the observations are shown together with the predictive densities given by Model 3, with warmer colors having higher likelihood. Notice that practically all the observations are covered by the conditional densities. Another feature of this model is that it assigns zero probability to events outside the state space, that is, for values of irradiance higher than the maximum or lower than zero. Comparatively, the predictive density in the bottom plot of Figure 3 is more spread out, as here we represent predictions issued 24 h ahead instead of 1 h ahead. A further illustration of this is seen in Figure 4, where the 95% prediction interval is shaded in gray.

Clearly, the prediction interval becomes skewed towards the center as we approach the limits of the state space, that is, when the process comes closer to the maximum irradiance or to zero. Upon careful inspection, it can be found that the prediction interval is the widest when we predict around 50%, which is to be expected from the physics of the system. Besides, as result of the model, the 24-h ahead forecast has a wider 95% prediction interval than that of the forecast issued 1-h ahead.

To validate the accuracy of the predictive density, we can evaluate the predictive quantiles of the distribution. This is carried out by counting how many observations lie on each side of the predictive quantile in question and comparing it with the expected number.

The exceedances of the predictive quantiles for Model 3 are shown in Table 3. The predictive distribution is found by simulating the process via Monte Carlo simulation of 1000 sample trajectories. In a perfect data fit, the expected quantiles match the observed ones exactly. For the 1-h prediction on the training set, we see an excellent performance, with the frequency of exceedances quite close to the expected one in a perfect fit. For the 24-h prediction horizon, we see a slightly lower number of exceedances than expected. This is also true for the

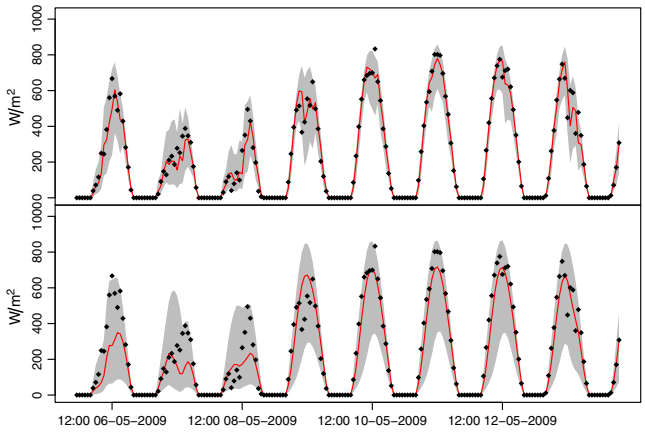

Figure 4. Observations of solar irradiance in black plotted along with the 1-h ahead (top) and the 24-h ahead (bottom) 95% prediction intervals in gray and the prediction in red

Table 3. Frequency of observed exceedances for selected quantiles of the predictive density given by the quantile function $Q(\cdot)$

		Training set		Test set	
Quantile function	Expected	1h	24h	1h	24h
$Q(0.1)$	0.10	0.088	0.079	0.076	0.061
$Q(0.2)$	0.20	0.176	0.170	0.156	0.141
$Q(0.3)$	0.30	0.273	0.261	0.253	0.220
$Q(0.4)$	0.40	0.376	0.348	0.349	0.301
$Q(0.5)$	0.50	0.486	0.440	0.458	0.392
$Q(0.6)$	0.60	0.603	0.540	0.589	0.473
$Q(0.7)$	0.70	0.720	0.645	0.712	0.580
$Q(0.8)$	0.80	0.818	0.763	0.811	0.728
$Q(0.9)$	0.90	0.901	0.885	0.902	0.858

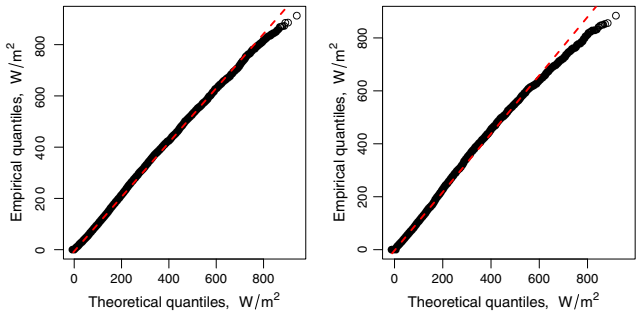

Figure 5. Q–Q plots for the training set (left) and the test set (right) based on the empirical and simulated data from the model over the respective period

test set, especially for the 24-h ahead quantiles. It should be pointed out here that Model 3 has been fitted on the basis of one-step-ahead predictions, that is, the prediction of the next hour on the training set. Thus, the model is not tuned to predictions for a 24-h horizon, even though it seems to perform reasonably well.

To evaluate how well Model 3 captures the long-term (stationary) behavior of the historical series of solar irradiance, we compare the empirical quantiles of the data with quantiles generated by simulating the model. This is carried out in Figure 5, where it can be seen that the empirical and theoretical quantiles lie on a somewhat straight line. Furthermore, the performance on the training set is better than on the test set, as one would expect a priori. The stationary distribution of the model on the test set, nevertheless, seems to be satisfactorily aligned with the empirical distribution.

5. CONCLUDING REMARKS

With the increasing penetration of renewable generation in energy systems, forecasting renewable production is becoming crucial for its efficient integration. The optimal solution to decision-making problems characterized by asymmetric costs in time requires an understanding of the uncertainty associated with the stochastic renewable production. This is the case, for example, of the decision-making process faced by a stochastic renewable producer that desires to trade its production in an electricity market such as the Nordic Power Exchange (Nordpool), where power is sold 12–36 h ahead of its physical delivery. If the renewable power producer is then not able to fulfill its forward contract obligations, it is required to procure the missing power in the balancing (real time) market, where it will be charged an energy price that may depend on the sign of her production imbalance. This paper proposes an SDE framework for modeling the uncertainty associated with solar irradiance. It allows us to capture the dynamics of a stochastic process that is confined to a bounded state space by specifying an SDE model that naturally satisfies those physical restrictions. This is very useful for probabilistic forecasting. Needless to say, this could also be indirectly achieved by means of truncated distributions. However, the world of SDEs provides us with a flexible framework to produce models for bounded stochastic processes with a better performance and characterized just by a few parameters.

The methodological approach followed in this paper is likelihood estimation based on second-order moment representation (through the extended Kalman filter) of the solar irradiance dynamics. Clearly, other approaches could be used in this context, in particular, simulation-based methods (Dowd (2007)) and direct estimation of the Fokker–Planck equation (Pedersen *et al.* (2008)). In both cases, the major drawback is the computational cost of evaluating the likelihood. More specifically, in the case of simulation-based methods, the computational cost is related to the number of data points (Dowd (2007) use 190 datapoint), whereas for the direct evaluation of the Fokker–Planck equation, the computational cost is related to the complexity of the SDE (Pedersen *et al.* (2008) use a very simple SDE with complicated boundary conditions). In our view, the combination of the extended Kalman filter and the Lamperti transform offers a flexible framework for SDE estimation with a relatively large number of data points. Furthermore, the in-sample and out-of-sample results reported in Section 4 indicate that the linearization introduced by the filter performs satisfactorily.

The starting point for the modeling carried out in this paper is a simple SDE that tracks the expected solar irradiance from an NWP. By normalizing the weather prediction with the maximum irradiance, we can capture the periodic behavior in the process dynamics, and consequently, achieve major improvements. We can tune the diffusion term to model the actual behavior of the process and confine it to a bounded interval. The SDE formulation allows for formulating complex model structures and to track conditional distributions at any point in time. Our proposed SDE modeling approach outperforms simple as well as more complex benchmarks.

Even though there is a relation between solar irradiance and produced power from a photovoltaic panel, such a relation is not trivial. It depends on tilt and orientation of the photovoltaic (PV) panel as well as on its efficiency, which may vary as the panel becomes dirty or deteriorates over time. To address this, an adaptive estimation approach would be appropriate. Future studies will be directed at constructing models to capture this and produce probabilistic forecasts for solar power via a power curve. Because what is important to the energy system is the total input of renewable energy, future studies will also be directed at comodeling wind and solar power, as these are expected to be

main contributors to the energy mix of the future. Another potential line of future research is the modeling of the interdependence between the power output of solar farms at different locations.

REFERENCES

Adomian G. 1988. *Nonlinear Stochastic Systems Theory and Application to Physics*, Vol. 46. Springer: Dordrecht, The Netherlands.

Arulampalam MS, Maskell S. 2002. Gordon N. *IEEE Transactions on Signal Processing* **50**(2):174–188.

Bacher P, Madsen H, Nielsen H. 2009. Online short-term solar power forecasting. *Solar Energy* **83**(10):1772–1783.

Bacher P, Madsen H, Perers B, Nielsen H. 2013. A non-parametric method for correction of global radiation observations. *Solar Energy* **88**:13–22.

Bera AK, Bilias Y. 2002. The MM, ME, ML, EL, EF and GMM approaches to estimation: A synthesis. *Journal of Econometrics* **107**(1):51–86.

Bhardwaj S, Sharma V, Srivastava S, Sastry OS, Bandyopadhyay B, Chandel SS, Gupta JRP. 2013. Estimation of solar radiation using a combination of Hidden Markov Model and generalized Fuzzy model. *Solar Energy* **93**:43–54.

Bird RE, Hulstrom RL. 1981. Simplified clear sky model for direct and diffuse insolation on horizontal surfaces. *Technical Report*, Solar Energy Research Inst. Golden, CO (USA).

Björk T. 2009. *Arbitrage Theory in Continuous Time*. OUP Oxford: Oxford Finance Series.

Boland J. 2008. Time series modelling of solar radiation. In *Modeling Solar Radiation at the Earth's Surface,* Badescu V (ed.) Springer: Berlin Heidelberg; 283–312.

Box GE, Cox DR. 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* **26**(2):211–252.

Breinholt A, Thordarson FÖ, Møller JK, Grum M, Mikkelsen PS, Madsen H. 2011. Grey-box modelling of flow in sewer systems with state-dependent diffusion. *Environmetrics* **22**(8):946–961.

Chen C, Duan S, Cai T, Liu B. 2011. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Solar Energy* **85**(11):2856–2870.

Chui CK, Chen G. 2009. *Kalman Filtering: With Real-Time Applications*. Springer: Berlin Heidelberg.

Dowd M. 2007. Bayesian statistical data assimilation for ecosystem models using Markov chain Monte Carlo. *Journal of Marine Systems* **68**(3):439–456.

European Photovoltaic Industry Association. 2013. Global market outlook for photovoltaics 2013–2017. Online, http://www.epia.org/news/publications/ [Accessed 9 December 2013].

Gneiting T, Raftery A. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**(477):359–378.

Hacker JP, Rife DL. 2007. A practical approach to sequential estimation of systematic error on near-surface mesoscale grids. *Weather and Forecasting* **22**(6):1257–1273.

Huang J, Korolkiewicz M, Agrawal M, Boland J. 2013. Forecasting solar radiation on an hourly time scale using a Coupled AutoRegressive and Dynamical System (CARDS) model. *Solar Energy* **87**(1):136–149.

Ineichen P. 2006. Comparison of eight clear sky broadband models against 16 independent data banks. *Solar Energy* **80**(4):468–478.

Jazwinski AH. 2007. *Stochastic Processes and Filtering Theory*. Courier Dover Publications: Mineola, New York.

Ji W, Chee KC. 2011. Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. *Solar Energy* **85**(5):808–817.

Juhl R, Kristensen NR, Bacher P, Kloppenborg J, Madsen H. 2013. CTSM-R user guide. *Technical University of Denmark*.

Kaplanis S, Kaplani E. 2010. Stochastic prediction of hourly global solar radiation for Patra, Greece. *Applied Energy* **87**(12):3748–3758.

Kasten F, Young AT. 1989. Revised optical air mass tables and approximation formula. *Applied Optics* **28**(22):4735–4738.

Kloeden PE, Pearson R. 1977. The numerical solution of stochastic differential equations. *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics* **20**(01):8–12.

Lorenz E, Hurka J, Heinemann D, Beyer HG. 2009. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2**(1):2–10.

Madsen H, Holst J. 1995. Estimation of continuous-time models for the heat dynamics of a building. *Energy and Buildings* **22**(1):67–79.

Madsen H, Thyregod P. 2011. *Introduction to General and Generalized Linear Models*. Chapman & Hall/CRC Texts in Statistical Science Series, CRC PressINC: Boca Raton, Florida, USA.

Mathiesen P, Brown J, Kleissl J. 2013. Geostrophic wind dependent probabilistic irradiance forecasts for coastal california. *IEEE Transactions on Sustainable Energy* **4**(2):510–518.

Mihalakakou G, Santamouris M, Asimakopoulos D. 2000. The total solar radiation time series simulation in Athens, using neural networks. *Theoretical and Applied Climatology* **66**(3–4):185–197.

Mikosch T. 1998. Elementary Stochastic Calculus with Finance in View **6**.

Møller J, Madsen H. 2010. From state dependent diffusion to constant diffusion in stochastic differential equations by the Lamperti transform. *Technical Report*, Technical University of Denmark.

Møller J, Pinson P, Madsen H. 2013. Probabilistic forecasts of wind power generation by stochastic differential equation models. *Technical Report*, Technical University of Denmark.

Møller JK, Madsen H, Carstensen J. 2011. Parameter estimation in a simple stochastic differential equation for phytoplankton modelling. *Ecological Modelling* **222**(11):1793–1799.

Nicolau J. 2002. A new technique for simulating the likelihood of stochastic differential equations. *The Econometrics Journal* **5**(1):91–103.

Øksendal B. 2010. *Stochastic Differential Equations: An Introduction with Applications*. Universitext (1979), Springer: London, United Kingdom.

Pedersen MW, Righton D, Thygesen UH, Andersen KH, Madsen H. 2008. Geolocation of North Sea cod (Gadus morhua) using hidden Markov models and behavioural switching. *Canadian Journal of Fisheries and Aquatic Sciences* **65**(11):2367–2377.

Pedro H, Coimbra C. 2012. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy* **86**(7):2017–2028.

Perez R, Kivalov S, Schlemmer J, Hemker K, Jr, Renné D, Hoff TE. 2010. Validation of short and medium term operational solar radiation forecasts in the US. *Solar Energy* **84**(12):2161–2172.

Pinson P, Nielsen H, Møller J, Madsen H, Kariniotakis G. 2007. Non-parametric probabilistic forecasts of wind power: Required properties and evaluation. *Wind Energy* **10**(6):497–516.

Ridley B, Boland J, Lauret P. 2010. Modelling of diffuse solar fraction with multiple predictors. *Renewable Energy* **35**(2):478–483.

Ristic B, Arulampalm S, Gordon NJ. 2004. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House Publishers: London, United Kingdom.

Rousseeuw PJ, Leroy AM. 2005. *Robust Regression and Outlier Detection*, Vol. 589. Wiley. com: Hoboken, New Jersey, USA.

Soubdhan T, Emilion R. 2010. Stochastic differential equation for modeling global solar radiation sequences. *Proceedings of the IASTED International Conference: Modelling, Identification, and Control (AsiaMIC 2010)*: Phuket, Thailand, 14–17.

Van Kampen NG. 1992. *Stochastic Processes in Physics and Chemistry*, Vol. 1. Access Online via Elsevier: Amsterdam, The Netherlands.
Yang D, Jirutitijaroen P, Walsh W. 2012. Hourly solar irradiance time series forecasting using cloud cover index. *Solar Energy* **86**(12):3531–3543.
Zhou Z, Botterud A, Wang J, Bessa R, Keko H, Sumaili J, Miranda V. 2013. Application of probabilistic wind power forecasting in electricity markets. *Wind Energy* **16**(3):321–338.

Paper D

Short-term Probabilistic Forecasting of Wind Speed using Stochastic Differential Equations

Accepted in *International Journal of Forecasting, 2014*.

Short-term Probabilistic Forecasting of Wind Speed Using Stochastic Differential Equations

Emil B. Iversen, Juan M. Morales, Jan K. Møller, Henrik Madsen

Technical University of Denmark, Matematiktorvet, building 303b, DK-2800 Lyngby, Denmark

Abstract

It is widely accepted today that probabilistic forecasts of wind power production constitute valuable information for both wind power producers and power system operators to economically exploit this form of renewable energy, while mitigating the potential adverse effects related to its variable and uncertain nature. To provide reliable wind power forecasts ranging beyond a couple of hours, forecasts of the wind speed are fundamental. In this paper, we propose a modeling framework for wind speed that is based on stochastic differential equations. We show that stochastic differential equations allow us to naturally capture the time dependence structure of wind speed prediction errors (from 1 up to 24 hours ahead) and, most importantly, to derive point and quantile forecasts, predictive distributions, and time-path trajectories (also referred to as scenarios or ensemble forecasts), all by one single stochastic differential equation model characterized by a few parameters.

Keywords: Wind speed, Probabilistic Forecasting, Wind Power, Stochastic Differential Equations

1. Introduction

The last few years have witnessed a remarkable increase in the contribution of renewable energy sources to the global electricity supply, with the largest share coming from wind turbines in many countries (The European Wind Energy Association (2013)). Wind power production is, however, highly variable and uncertain, thus challenging the traditional practices for power system operation and the trade of this form of renewable energy in electricity markets. In order to mitigate the adverse effects of the stochastic

nature of wind, good forecasts of the power generated by wind farms is a must. Furthermore, to fully manage and exploit wind energy, these forecasts should not only provide power system operators and wind power producers with a single-valued guess of the future wind generation—a so-called *point forecast*—but also with information on possible outcomes and their associated probability of occurrence. This enriched form of forecasting is known as *probabilistic forecasting* and takes on its full meaning in the context of wind power management and trading (Morales et al. (2014); Zhou et al. (2013)).

The literature on wind power forecasting is now vast, but mostly centered on techniques for point predictions. For a comprehensive review of the topic, the interested reader is referred to Costa et al. (2008); Monteiro et al. (2009); Foley et al. (2012). Methods for wind power probabilistic forecasting are, on the contrary, not so developed. Wind power density forecasts are commonly obtained by superimposing a model for the probability distribution of prediction errors on a point forecast (typically, the average or most likely outcome) (Bremnes, 2006; Møller et al., 2008; Pinson and Kariniotakis, 2010), or by post-processing ensemble forecasts from meteorological models so that they represent the true predictive density (Nielsen et al., 2006; Pinson and Madsen, 2009; Taylor et al., 2009). In the realm of probabilistic forecasting, it is customary to distinguish between *parametric* and *nonparametric* methods. The former presuppose a certain standard distribution for the forecast error and as a result, the modeling endeavor boils down to estimating the parameters characterizing such a distribution (Pinson, 2012; Messner et al., 2013; Thorarinsdottir and Gneiting, 2010; Thorarinsdottir and Johnson, 2012). In contrast, the non-parametric methods do not assume any pre-specified forecast error distribution and work directly with the empirical distribution instead (Messner et al., 2013; Bessa et al., 2012; Pinson et al., 2007; Bremnes, 2004). Recent works have also focused on generating realistic sample trajectories of the stochastic process and producing multi-horizon or multi-variate probabilistic forecasts. To this end, the most popular approach is to fit a marginal predictive density for each univariate output variable (e.g., the wind speed in each time period of the prediction horizon)—see Lerch and Thorarinsdottir (2013) for a comparison of different regression-type models for wind speed—and then combine these marginals into a multivariate cumulative density using copula theory (Schefzik et al., 2013).

Forecasts of wind power are markedly improved by the use of numerical weather predictions (NWPs) of wind speed. Foley et al. (2012) provide an

overview of the different uses of NWPs for wind power forecasting. Specifically, they underline that ensemble forecasts of wind speed can be exploited to obtain valuable information on the reliability of the wind power forecast. The usefulness of NWPs of wind speed for wind power forecasting has also been stressed, for instance, in Ramirez-Rosado et al. (2009) and De Giorgi et al. (2011). In Taylor et al. (2009) and Pinson and Madsen (2009), predictive densities of wind speeds are used to produce probabilistic forecasts of wind power. These predictive densities are estimated from the ensemble forecasts provided by a NWP system.

This paper describes a novel approach to wind speed probabilistic forecasting based on stochastic differential equations (SDEs). The proposed SDE model upgrades the numerical weather prediction by using wind speed data and provides plausible time-path trajectories of the wind speed process that are perfectly comparable to the ensemble forecasts obtained from a NWP system. Furthermore, our SDE model can generate these trajectories for a specific wind site swiftly. For these reasons, our SDE model can significantly contribute to wind power forecasting. More generally, SDEs offer a powerful and versatile modeling framework that allows us to consistently issue point- and all forms of probabilistic forecasts (namely, quantiles, densities, or time-path trajectories) by the same model. Moreover, the parameters characterizing the SDE can be intuitively interpreted, which makes it much easier to formulate model extensions on the basis of the specific physics of the underlying stochastic process or of observable statistical deficiencies. The proposed modeling framework naturally captures the time dependence of forecast errors and events with zero probability, such as negative wind speeds, and does not need to assume Gaussian innovations. Seen in a broader perspective, SDEs cover the large class of stochastic processes with continuous trajectories and, in fact, many of the discrete time models used in classical time series theory can be seen as discrete-time versions of SDEs.

The application of stochastic differential equations to forecasting and, in particular, to wind power forecasting is a very recent topic and consequently, the technical literature in this regard is scant. Two works, however, should be mentioned here, namely, Møller et al. (2013) and Zárate-Miñano et al. (2013). SDEs are fruitfully used in Møller et al. (2013) for wind power forecasting by considering state-dependent diffusions and a numerical weather prediction as external input. The SDE model proposed by the authors is simultaneously fitted to data ranging from 1 to 48 hours ahead, which makes it computation-

ally intensive to estimate and limits the amount of data that can be used for fitting. The SDE model proposed in our paper is estimated, on the contrary, on one-step-ahead data and thus, the approach that we use here resembles that of the Box-Jenkins type models that are fitted using one-step-ahead information. Furthermore, our model focuses on wind speed, and not on wind power, and therefore, the challenges are different. In Zárate-Miñano et al. (2013), a continuous time model for wind speed is presented. It is intended to simulate wind speed trajectories over very short time horizons. It cannot be used for forecasting, though, as parameters are not estimated, no external inputs are allowed, and the SDE model is limited to a very simple structure. In addition, the SDE model the authors in Zárate-Miñano et al. (2013) propose is designed to fit the long-term stationary distribution of wind speed. The stationary distribution, however, does not necessarily hold for the short term, as one can see for the case of the climatological forecast.

The approach taken to modeling wind speed in this paper is similar to the approach adopted to modeling solar irradiance in Iversen et al. (2014) in that they both rely on SDEs. However, the weather phenomena considered in these two papers, namely, wind speed and solar irradiance, are remarkably different, each with its own challenges. Indeed, the modeling of wind speed substantially differs from that of solar irradiance in terms of the physical domain of the underlying process and its periodic nature. This results in distinct structures for the drift and diffusion terms. Moreover, in the case of wind speed, the mere introduction of the NWP as an exogenous input to the SDE model is not advantageous, as it results in the simulated process systematically lagging behind the NWP. This will be shown later on. We solve this issue by introducing the derivative of the NWP as well. Furthermore, in the present paper, we provide a straightforward methodology to generate predictive densities and time-path trajectories of wind speeds analogous to the ensemble forecasts obtained from a NWP system. These ensemble forecasts are of particular relevance for wind power forecasting, whereas time-path trajectories and predictive densities for solar irradiance are still today of limited use, at least comparatively speaking.

The remainder of the paper is organized as follows: Section 2 provides a short introduction to stochastic differential equations (SDEs) and the parameter estimation procedure. Section 3 provides a model for wind speeds based on stochastic differential equations. This model is subsequently validated on a training and test set in Section 4. Section 5 concludes the paper and provides directions for future research.

2. Stochastic Differential Equations

A stochastic differential equation (SDE) is a differential equation with one or more stochastic terms resulting in a solution that is in itself a stochastic process. SDEs are used to describe various phenomena driven by a large random component and are especially prominent in mathematical finance (Björk, 2009; Mikosch, 1998) and physics (Van Kampen, 1992; Adomian, 1988). We give here a very short introduction to SDEs and refer the interested reader to Øksendal (2010) for a thorough and mathematically rigorous discussion on the topic.

A SDE is commonly stated as

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t, \tag{1}$$

where W_t is a Wiener process and functions $f(\cdot)$ and $g(\cdot)$ are known as the drift and diffusion terms, respectively. However, Equation (1) is not well defined, as the derivative of W_t, dW_t, does not exist. Thus, Equation (1) should actually be seen as a shorthand for the integral equation

$$X_t = X_0 + \int_0^t f(X_s, s)ds + \int_0^t g(X_s, s,)dW_s. \tag{2}$$

We use here the Itō interpretation of the second integral.

In contrast to the solution of an ordinary differential equation, which consists of a single time trajectory representing the value of the modeled process at each point in time into the future, the solution to a SDE is a stochastic process, which characterizes the uncertainty in the system dynamics for each future time. For an Itō process given by the SDE in (1) with drift $f(X_t, t)$ and diffusion coefficient $g(X_t, t) = \sqrt{2D(X_t, t)}$, the density $j(x, t)$ of the random variable X_t in state x at time t is the solution to the partial differential equation (Björk, 2009):

$$\frac{\partial}{\partial t} j(x, t) = -\frac{\partial}{\partial x} \left[f(x, t) j(x, t) \right] + \frac{\partial^2}{\partial x^2} \left[D(x, t) j(x, t) \right], \tag{3}$$

which is known as the Fokker-Planck equation or the Kolmogorov forward equation. Hence, given a specific SDE, the probability density function of the process can be found by solving a partial differential equation (PDE). In general, this PDE cannot be solved analytically, but, fortunately, a wide range of numerical solution approaches do exist.

It can be shown (see the Lévy-Itō decomposition in Björk (2009)) that, under sufficient regularity conditions, all stochastic processes with continuous trajectories can be written as special cases of SDEs. Therefore, SDEs are a general class of stochastic processes. Indeed, many ordinary time series models can be interpreted as discrete versions of SDEs.

In practice, it is only possible to observe continuous-time systems in discrete time. For this reason, one defines the observation Y_k of the process at time t_k, which is found through some measurement equation $h(\cdot)$. We thus have the following system of equations:

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t \tag{4}$$
$$Y_k = h(X_{t_k}, t_k, e_k), \tag{5}$$

where e_k is some measurement error.

The procedure that we use in this paper to estimate (4) and (5) relies on specific forms of $g(\cdot)$ and $h(\cdot)$. In particular, we require that $g(X_t, t) = g(t)$ and $h(X_{t_k}, t_k, e_k) = h(X_{t_k}, t_k) + e_k$, where $e_k \sim \mathcal{N}(0, \sigma^2)$. In principle, even though these two constraints may considerably limit the modeling capability of our SDE framework, they can actually be relaxed to a large extent. Indeed, the first condition, which establishes that the diffusion term must not depend on the current state X_t, can be overcome by transforming the original SDE into an equivalent one with non-state-dependent diffusion using Itō calculus and the so-called *Lamperti transform* (Møller et al., 2008; Iversen et al., 2014). The second condition, which requires the observation error be additive and Gaussian, can be mitigated by transformations of the data (Box and Cox, 1964).

The procedure to estimate the vector θ of the parameters defining the SDE model (4)–(5) is based on the extended Kalman filter (see, for example, Welch and Bishop (1995)). The filter is applied to the equivalent SDE system that results from the aforementioned transformations. The first step in the estimation procedure is to find the one-step predictions of the mean and variance of the observations, which are defined as

$$\widehat{Y}_{k|k-1} = \mathbb{E}\left[Y_k | \mathcal{Y}_{k-1}, \theta\right] \tag{6}$$
$$R_{k|k-1} = \mathbb{V}\left[Y_k | \mathcal{Y}_{k-1}, \theta\right], \tag{7}$$

where $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ denote the expectation and variance, respectively, and $\mathcal{Y}_{k-1} = \{Y_0, \ldots, Y_{k-1}\}$. We use the extended Kalman filter (Jazwinski, 2007)

to determine these predictions. We can now define the innovation

$$\epsilon_k = Y_k - \widehat{Y}_{k|k-1}, \tag{8}$$

in order to compute the likelihood. For a system satisfying the conditions on $g(\cdot)$ and $h(\cdot)$ that we stated above, the approximated likelihood is given by

$$L\left(\theta; \mathcal{Y}_N\right) = \left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\epsilon_k^\top R_{k|k-1}^{-1}\epsilon_k\right)}{\sqrt{\det\left(R_{k|k-1}\right)}\left(\sqrt{2\pi}\right)^l} \right) p(Y_0|\theta), \tag{9}$$

where l is the dimension of the sample space, that is, the dimension of Y_k, N is the number of observations, $(\cdot)^\top$ denotes the vector transpose and $p(Y_0|\theta)$ is the likelihood of seeing observation Y_0. We cannot compute this likelihood, because there are no observations previous to Y_0. To get around this issue, we optimize instead the logarithm of the likelihood function conditional on Y_0, which results in

$$\begin{aligned} \log\left(L\left(\theta; \mathcal{Y}_N | \, Y_0\right)\right) \;=\; & -\frac{1}{2}\sum_{k=1}^{N}\left(\log(\det(R_{k|k-1})) + \epsilon_k^\top R_{k|k-1}^{-1}\epsilon_k\right) \\ & - \log(2\pi)\frac{Nl}{2}. \end{aligned} \tag{10}$$

The parameter vector θ enters the log-likelihood function (10) through ϵ_k and $R_{k|k-1}$. This parameter vector can now be estimated by maximizing (10), i.e.,

$$\hat{\theta} = \arg\max_{\theta\in\Theta}\left(\log\left(L\left(\theta; \mathcal{Y}_N | \, Y_0\right)\right)\right), \tag{11}$$

where Θ is the feasible parameter space. A thorough introduction to parameter estimation and filtering is found in Jazwinski (2007) and a more detailed description of the exact implementation of the extended Kalman filter and the parameter estimation procedure employed here can be found in Kristensen et al. (2004). We note that the likelihood function is optimized for the one-step-ahead residuals. To estimate SDE models using a multi-horizon approach, we refer the interested reader to Møller et al. (2013).

3. A SDE Model for Wind Speed

In this section, we present a probabilistic model for wind speed. This section follows the general model-building approach presented in Iversen et al.

(2014), which essentially applies general guidelines for system identification (see, e.g., Madsen (2008) and Ljung (1999)) to the context of stochastic differential equations. We start by introducing a basic SDE system that aims to track a given numerical weather prediction. We then highlight the shortcomings in this simple SDE model and use them as the basis to justify the different structural components of the more complex SDE model that we propose.

The data used in this study belong to a meteorological station located in the western part of Denmark and includes hourly wind-speed measurements together with predicted wind speeds based on a numerical weather prediction model from the Danish Meteorological Institute (Källén, 1996). The numerical weather prediction (NWP) provides a 48-hour forecast of the wind speed and is updated every 6 hours. The data cover three years, from 01/01-2009 to 31/12-2011 and is divided into two periods: the training set of two years length for parameter estimation and a test set spanning the remaining one year for evaluating the performance of the proposed SDE model.

We use the simple SDE model (12)–(13) below as a starting point.

$$dX_t = \theta_x(p_t\mu_x - X_t)dt + \sigma_x dW_t \tag{12}$$

$$Y_k = X_{t_k} + \epsilon_k. \tag{13}$$

Here we aim to track the NWP supplied by the Danish Meteorological Institute. In Equations (12) and (13), wind speed observations are given by Y_k. We let p_t denote the numerical weather prediction at time t. Parameter μ_x is a local scaling of the numerical weather prediction, which corrects for the NWP either over- or undershooting on average. $\theta_x \geq 0$ is a time constant, governing how rapidly the model returns to the predicted wind speed. Essentially, this parameter controls the relative contribution of the past observations and the NWP to the future wind speed. Indeed, a large θ_x results in a stochastic process that is mainly driven by the NWP (the process moves fast towards the NWP), while a small θ_x renders a stochastic process that is predominantly governed by the past observations of wind speed (the process moves slowly away from the previous wind speed value). Parameter σ_x characterizes the system noise. The data that we use are average hourly wind speeds measured by an anemometer. We include the stochastic variable $\epsilon_k \sim \mathcal{N}(0, \sigma_\epsilon)$ in the observation equation to characterize the measurement error of the physical system.

SDE model (12)–(13), however simple, suffers from grave deficiencies. First, the model assigns a positive probability to negative wind speeds, as

can be seen in Figure 1, where the one-step ahead (i.e., one-hour ahead) predictive distribution of wind speed is shown. This is clearly unrealistic. We will correct this obvious shortcoming by means of a state-dependent diffusion term.

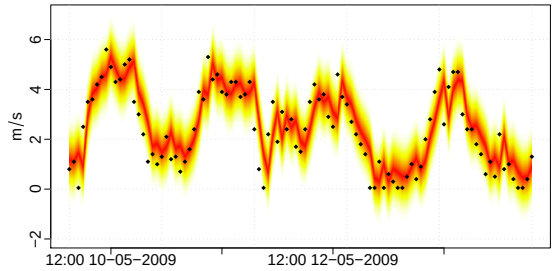

Figure 1: *One-step ahead predictive density from the model specified by Equations (12) and (13), approximated by Monte Carlo simulation. Warmer colors represents higher probability of seeing this realization. The black dots are the actual wind speed observations.*

Furthermore, the point forecast provided by model (12)–(13) for horizons longer than 1-hour is systematically shifted in time with respect to the numerical weather prediction (see Figure 2). We will resolve this issue by introducing the time derivative of the NWP as input to our SDE model.

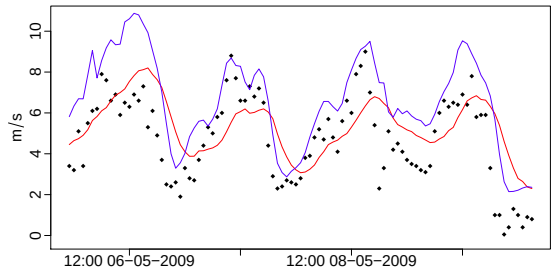

Figure 2: *The 24-hour ahead point prediction obtained from the model defined by Equations (12)–(13) in red. The NWP is shown in blue along with the observations in black.*

Finally, the autocorrelation function for the studentized residuals resulting from model (12)–(13), shown in Figure 3, reveals explicable structures in the wind speed process (e.g., a daily pattern) that are not being captured by this simple SDE model. Notice that, in Figure 3, the y-axis has been scaled down to better show the significance bounds. We will address this issue by allowing for a diurnal variation and a stochastic trend in the SDE system.

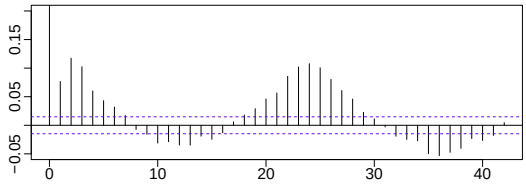

Figure 3: *The autocorrelation function for the studentized residuals of the SDE model defined by Equations (12)–(13).*

We propose the following SDE model for wind speeds:

$$\begin{aligned} dX_t &= \left(\left(\alpha_x \sin\left(\frac{2\pi}{24}t + \phi_x\right) + U_t + \rho_x \dot{p}_t\right)\left(1 - e^{-X_t}\right)\right. \\ &\quad \left. + \theta_x(p_t\mu_x - X_t)\right)dt + \sigma_x X_t^{\beta_x} dW_{x,t} \qquad (14) \\ dU_t &= \theta_u\left(\mu_u - U_t\right)dt + \sigma_u dW_{u,t} \qquad (15) \\ Y_k &= X_{t_k} + \epsilon_k. \qquad (16) \end{aligned}$$

Model (14)–(16) is characterized by the state-dependent diffusion term $\sigma_x X_t^{\beta_x} dW_{x,t}$, with parameter $\beta_x \geq 0$ determining its shape. Furthermore, as the process X_t tends to zero, so does the diffusion term. Consequently, within the vicinity of null wind speeds, the stochastic process X_t is dominated by the drift term, which will eventually push the process away from zero. Indeed, under certain regularity conditions, related to the relative size between the diffusion term and the drift term, and given that p_t is always larger than zero, we find that X_t has zero probability of taking on negative values. Notice that the term $\left(1 - e^{-X_t}\right)$ guarantees that $\left(\alpha_x \sin\left(\frac{2\pi}{24}t + \phi_x\right) + U_t + \rho_x \dot{p}_t\right)$ does not force the wind speed process out of the domain $\mathbb{R}^+$, since $\left(1 - e^{-X_t}\right)$ tends to zero as X_t approaches zero.

In order to remove the time shift between the numerical weather prediction and the point forecast of the SDE model, the time derivative of p_t, i.e., $\dot{p}_t$, is introduced. Similar to μ_x, ρ_x is a factor that scales $\dot{p}_t$. It is important to stress that model (14)–(16) is still causal, because $\dot{p}_t$, just like p_t, constitutes information that is available at time t as a by-product of the numerical weather prediction model. Therefore, $\dot{p}_t$ can be used as an input to predict the future evolution of the stochastic process Y_t. In practice, $\dot{p}_t$ should be seen as the *increment* of the wind speed process over the next time period that is forecast by the numerical weather prediction model, insomuch as the SDE model is specified in terms of increments.

Finally, U_t is a stochastic trend that captures the fact that periods of increasing (decreasing) wind speed tend to be followed by periods where the wind speed is also increasing (decreasing). In this line, μ_u describes the long-term trend of U_t and θ_u determines how rapidly U_t reverts to this level. The parameter σ_u governs the diffusion of U_t. Equation (14) includes a diurnal variation component, with α_x governing the amplitude and ϕ_x governing the phase. Both U_t and the diurnal variation are intended to improve on the NWP point forecast, which explains why they are introduced in the drift term.

4. Model Validation

We first show that the shortcomings in the basic SDE model (12)–(13) are no longer present in the proposed SDE model (14)–(16). In this vein, Figure 4 displays the predictive densities obtained from the latter. Note that, except for some small observation noise, negative wind speeds are now assigned zero probability.

Figure 5 shows the point prediction given by our SDE model and the associated 95%-prediction interval. The comparison between this figure and Figure 2 makes it clear that the time-shift problem highlighted in Section 3 is no longer an issue.

The autocorrelation function of the studentized residuals yielded by the proposed SDE model is depicted in Figure 6. The range on the y-axis has been scaled to $(-0.05, 0.20)$ to better illustrate the significance bands, which are quite small, because the data set is large. By comparing this figure with Figure 3, one can conclude that the structure around the lag 24 has almost disappeared, even though there are still some autocorrelations at a few lags

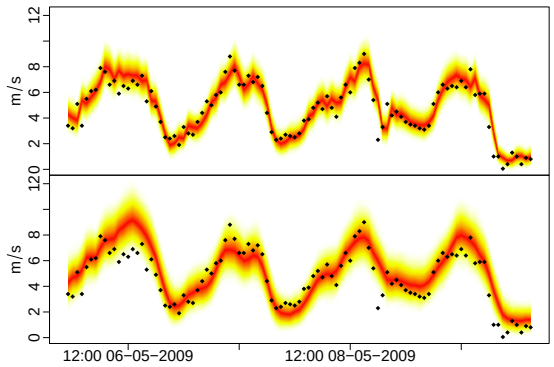

Figure 4: *1-hour (top) and 24-hour (bottom) ahead predictive density of the proposed SDE model, with warmer colors indicating a higher probability of seeing this realization. The densities are approximated by Monte Carlo simulations.*

that could be considered as statistically significant, indicating that the fit is, naturally, not completely perfect.

On the basis of these results, the proposed SDE model is deemed satisfactory on the training set. It only remains now to validate the model on an out-of-sample test set.

Since our aim is to provide the conditional distribution of the wind speeds at a future time, conventional evaluation methods, such as the mean absolute error (MAE) or the root mean squared error (RMSE) are not appropriate in this context, as they consider only the deviation from the point forecast. Instead, we evaluate the models in terms of likelihoods on the training and test sets and in terms of the continuous rank probability score (CRPS) on the test set. We compute the likelihood as described in Section 2, whereas the CRPS is calculated according to the definition provided in Gneiting and Raftery (2007). Both of these scores are proper scoring rules (Gneiting and Raftery, 2007), meaning that a better fit of the data will produce a better score.

To assess the performance of the proposed SDE model, we consider a number of benchmarks, namely, the climatological model, where we use the empirical density of the training set to predict the next observation; the

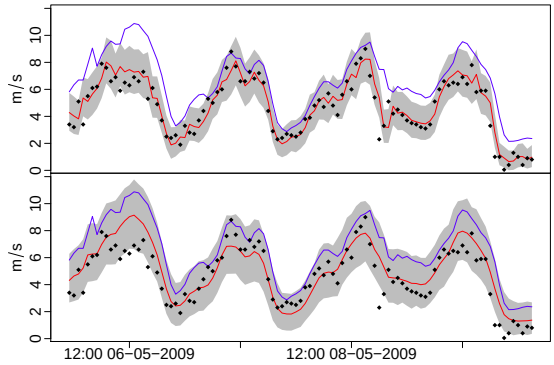

Figure 5: *1-hour (top) and 24-hour (bottom) ahead point prediction from the SDE model in red along with the 95%-prediction interval (gray shaded area). These are obtained through Monte Carlo simulations. The blue line is the point prediction of the NWP.*

persistence prediction, where the wind speed forecast is given by the present value; and a family of Auto-Regressive (AR) models. To be more precise, we consider a standard AR model and an ARX model, with the numerical weather prediction as exogenous input. This is then extended to have a truncated Gaussian innovation. Lastly, we consider an ARX-GARCH model, also with an extension to truncated Gaussian innovation. These benchmarks are classical time-series models (Madsen, 2008; Box et al., 2013; Hamilton, 1994). In particular, AR-GARCH methods have been used for forecasting daily wind speeds in Tol (1997). For shorter horizons, Box-Jenkins-type models have been used for modeling and forecasting wind speeds in Huang and Chalabi (1995) and Katz and Skaggs (1981), among others.

The persistance model is defined as:

$$Y_k = Y_{k-1} + \epsilon_k, \qquad \epsilon_{k+1} \sim \mathcal{N}(0, \sigma^2). \tag{17}$$

Here we have used a Gaussian innovation. While the persistence forecast does not require any associated distribution of the innovation, this is needed for computing a likelihood. We choose the innovation to be Gaussian to be in line with the classical AR model setup.

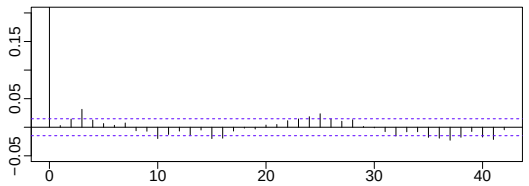

Figure 6: *The ACF of the proposed SDE model.*

The AR model is specified as:

$$Y_k = \psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \epsilon_k, \qquad \text{where } \epsilon_k \sim \mathcal{N}(0, \sigma^2). \tag{18}$$

The ARX model takes the form:

$$Y_k = \psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \phi p_k + \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}(0, \sigma^2). \tag{19}$$

For the ARX with the truncated Gaussian innovation, ϵ_k follows a Gaussian distribution truncated at $-\left(\psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \phi p_k\right)$. This way, the process is confined to $\mathbb{R}^+$.

The ARX-GARCH model is of the form:

$$Y_k = \psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \phi p_k + \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}(0, \sigma_k^2), \tag{20}$$

$$\sigma_k^2 = \alpha_0 + \sum_{i=1}^{\tilde{p}} \alpha_i \sigma_{k-i}^2 + \sum_{i=1}^{\tilde{q}} \beta_i \tilde{\epsilon}_{k-i} \qquad \tilde{\epsilon}_k \sim \mathcal{N}(0, \tilde{\sigma}^2), \tag{21}$$

where the Gaussian innovation, $\epsilon_k \sim \mathcal{N}(0, \sigma_k^2)$, is truncated at $-\left(\psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \phi p_k\right)$, for the variant of this benchmark with truncated innovation.

The scores for the SDE model and the benchmarks are indicated in Table 1. Clearly, the SDE model outperforms the benchmarks on the training and test sets both in terms of likelihood and CRPS. For the climatology benchmark, the empirical log-likelihood is used as it is a non-parametric

		Training Set	Test Set	
Models	Parameters	LL	LL	CRPS
Climatology	-	−36346	−19609	1.2577
Persistence	1	−23496	−12023	0.5223
AR	6	−22903	−11764	0.5098
ARX	7	−22087	−11332	0.4847
ARX - TN	8	−21445	−11012	0.4882
ARX - GARCH	9	−21715	−11081	0.4717
ARX - GARCH - TN	10	−21117	−10784	0.4734
SDE Model	11	**−20599**	**−10433**	**0.4468**

Table 1: *The log-likelihoods and CRPS of the proposed SDE model along with those of the selected benchmarks. The empirical log-likelihood is used for the climatology benchmark. The acronym TN stands for "truncated normal" (distribution).*

method, see for instance Bera and Bilias (2002). The SDE model provides thus significant improvements over simple as well as complex benchmarks, with only a slight increase in the number of parameters with respect to the benchmark with the best performance.

The parameter estimates of the fitted SDE model are collated in Table 2.

$\hat{\theta}_x$	0.425	$\hat{\alpha}_x$	0.269
$\hat{\mu}_x$	0.817	$\hat{\phi}_x$	−0.191
$\hat{\sigma}_x$	0.421	$\hat{\theta}_u$	0.0185
$\hat{\sigma}_\epsilon$	0.0623	$\hat{\mu}_u$	−0.0502
$\hat{\beta}_x$	0.423	$\hat{\sigma}_u$	0.0439
$\hat{\rho}_x$	0.904		

Table 2: *Parameter estimates for the SDE model.*

The values of these estimates are reasonable. For example, $\hat{\mu}_x$ and $\hat{\rho}_x$ in the SDE model are close to 1, indicating that the wind speed tends toward the numerical weather prediction, albeit with some bias.

The SDE model can also be used to generate multi-horizon predictive densities. These predictive densities can be approximated using Monte Carlo simulation and are shown in Figure 7, where we see that the density spreads out as the horizon increases. Figure 7 also includes a single simulated tra-

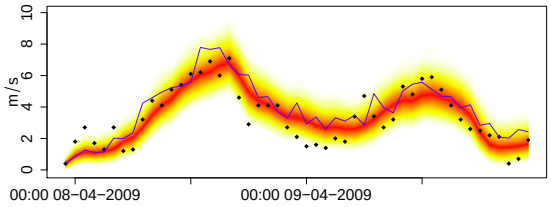

Figure 7: *The multi-step predictive densities of the SDE model for 1-48 hours ahead obtained via Monte Carlo simulation, with warmer colors indicating a higher probability. The blue line is a simulated plausible time-path trajectory (scenario) of the wind speed process and the black dots are the actual realized values.*

jectory of the wind speed process (scenario) along with its actual realization. Scenarios can be easily simulated from our SDE model by one of the many numerical approaches provided in Øksendal (2010) or Jazwinski (2007). Nonetheless, we limit ourselves here to evaluating the predictive densities of wind speed and thus, we do not go into the realm of verification methods for time-path trajectories. For a more detailed discussion on how to treat and evaluate time-path trajectories, we refer the interested reader to Pinson and Girard (2012) as a relevant practical case among recent developments of multivariate probabilistic forecast verification.

Models	Horizon			
	1 hour	4 hours	12 hours	24 hours
ARX - GARCH	0.4717	0.6168	0.6339	0.6357
ARX - GARCH - iterative	0.4717	0.6863	0.7464	0.7465
SDE model	0.4468	0.5713	0.6172	0.6236

Table 3: *The CRPS for the SDE model and the two ARX-GARCH benchmarks. The non-iterative ARX-GARCH is fitted specifically for each forecast horizon. The iterative ARX-GARCH model is fitted to 1-hour ahead data and then run iteratively until the desired prediction horizon is reached.*

As indicated in Section 2, the parameters in the SDE model are fitted to 1-hour ahead data. Nevertheless, we also investigate the performance of our SDE model for prediction horizons longer than one hour, in case that

our model were to be used for multi-horizon forecasting. For this purpose, the predictive densities obtained from the SDE model for four different lead times are compared against those given by an ARX-GARCH model with parameters fitted specifically for each lead time and those provided by an ARX-GARCH model with parameters fitted to 1-hour ahead data, which is then run iteratively until the targeted prediction horizon is reached. The comparison is carried out in terms of the CRPS. These ARX-GARCH models are employed as benchmarks here as they show the best performance from among all the benchmark models listed in Table 1. Notice that the proposed SDE model outperforms the two ARX-GARCH models. This is particularly remarkable, especially because one of the ARX-GARCH models is expressly fitted for the specific lead time under consideration. Furthermore, this ARX-GARCH model cannot be used to generate proper wind speed trajectories, while the proposed SDE model (and the "iterative" ARX-GARCH) can.

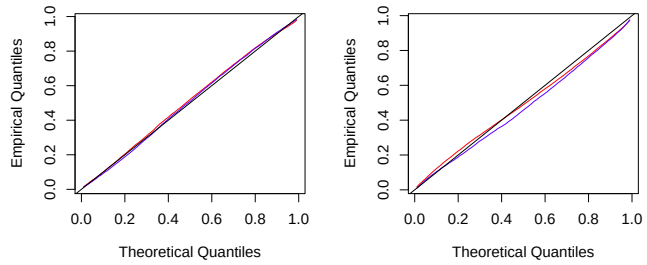

Figure 8: *QQ plot for the proposed SDE model for 1 hour ahead to the left and 24 hours ahead to the right. For each we show the quantiles for the training set in red and the quantiles for the test set in blue.*

To assess how the quantiles implied by the proposed SDE model match up with the empirical ones, Figure 8 shows two quantile-quantile plots corresponding to the 1-h and 24-h ahead predictions. These quantiles have been computed for both the training and the test sets, by using Monte Carlo simulation with 10 000 samples for each time step. It is apparent that the 1-h ahead quantiles agree accurately with the empirical ones, whereas the 24-

h ahead prediction underestimates the low quantiles and overestimates the larger quantiles slightly.

To conclude this section, we would like to mention that there exist other modeling approaches that are specially designed to perform multi-horizon forecasting. One of the most popular is to fit a marginal predictive density for each specific lead time and subsequently combine them into a multivariate cumulative density using copula theory (Schefzik et al., 2013; Bessa et al., 2012). Copula-based methods are able to capture a wide range of distributions and inter-dependencies. They are, however, computationally demanding and require the estimation of a high number of parameters. Stochastic differential equations can also be fit for multi-horizon forecasting (Møller et al., 2013). This approach relieves the issue of the large number of parameters to be estimated, but is still computationally very intensive and can only handle a limited amount of data. All that said, we find that our SDE approach, albeit based on a one-step-ahead estimation procedure, can also be considered as a reasonable method for multi-horizon forecasting given that it relies on relatively few parameters and demands a computational time comparable to standard time-series models, while showing a better performance.

5. Concluding Remarks

A modeling framework based on stochastic differential equations to describe wind speeds in continuous time is proposed. We build a model that captures the bounded nature of wind speed and its changing variability over time and that incorporates numerical weather predictions and stochastic trends. The model outperforms both simple and complex benchmarks on the training as well as the test set. The proposed modeling framework, based on SDEs, allows for easily formulating model extensions based on the physics of the system or statistical analysis. Among the different outputs of the model are accurate point forecasts, predictive densities, forecasts on multiple horizons which capture the interdependence in prediction errors, prediction intervals and scenarios. Because scenarios can be readily generated from the proposed SDE model using standard numerical techniques and because they preserve the same distribution as the one observed in the data, the generated scenarios would be able to replace ensemble forecasts, which are widely used as input to wind power forecasting.

Further research within the field of modeling energy systems using stochastic differential equations can be directed at co-modeling wind speed and solar

irradiance as these are the underlying physical phenomenona for two of the most rapidly developing renewable energy sources: wind and solar power. Aggregation of power production from larger areas can also be of great relevance. The framework of SDEs may also allow for spatio-temporal modeling of wind power by using stochastic partial differential equations. Another possibly fruitful topic for future research is to extend the proposed wind speed model to predict wind power, possibly by introducing a dynamic power curve.

Acknowledgments

DSF (Det Strategiske Forskningsråd) is to be acknowledged for partly funding the work of Emil B. Iversen, Juan M. Morales and Henrik Madsen through the Ensymora project (no. 10-093904/DSF). Furthermore, Juan M. Morales and Henrik Madsen are partly funded by the Research Centre CITIES (no. 1035-00027B), which is also supported by DSF (Det Strategiske Forskningsråd). Finally, we want to thank the three anonymous reviewers for their insightful and constructive comments, which have helped improve the paper significantly.

References

Adomian, G., 1988. Nonlinear Stochastic Systems Theory and Application to Physics. volume 46. Springer.

Bera, A.K., Bilias, Y., 2002. The MM, ME, ML, EL, EF and GMM approaches to estimation: A synthesis. Journal of Econometrics 107, 51–86.

Bessa, R.J., Miranda, V., Botterud, A., Zhou, Z., Wang, J., 2012. Time-adaptive quantile-copula for wind power probabilistic forecasting. Renewable Energy 40, 29–39.

Björk, T., 2009. Arbitrage Theory in Continuous Time. Oxford Finance Series, OUP Oxford. URL: `http://books.google.dk/books?id=N0ImZs3HpwIC`.

Box, G.E., Cox, D.R., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological) , 211–252.

Box, G.E., Jenkins, G.M., Reinsel, G.C., 2013. Time series analysis: forecasting and control. John Wiley & Sons.

Bremnes, J.B., 2004. Probabilistic wind power forecasts using local quantile regression. Wind Energy 7, 47–54.

Bremnes, J.B., 2006. A comparison of a few statistical models for making quantile wind power forecasts. Wind Energy 9, 3–11.

Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H., Feitosa, E., 2008. A review on the young history of the wind power short-term prediction. Renewable and Sustainable Energy Reviews 12, 1725–1744.

De Giorgi, M.G., Ficarella, A., Tarantino, M., 2011. Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods. Energy 36, 3968–3978.

Foley, A.M., Leahy, P.G., Marvuglia, A., McKeogh, E.J., 2012. Current methods and advances in forecasting of wind power generation. Renewable Energy 37, 1–8.

Gneiting, T., Raftery, A., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102, 359–378.

Hamilton, J.D., 1994. Time series analysis. volume 2. Princeton university press Princeton.

Huang, Z., Chalabi, Z., 1995. Use of time-series analysis to model and forecast wind speed. Journal of Wind Engineering and Industrial Aerodynamics 56, 311–322.

Iversen, E.B., Morales, J.M., Møller, J.K., Madsen, H., 2014. Probabilistic Forecasts of Solar Irradiance by Stochastic Differential Equations. Environmetrics (in press).

Jazwinski, A.H., 2007. Stochastic Processes and Filtering Theory. Courier Dover Publications.

Källén, E., 1996. HIRLAM documentation manual; 1996. System 2.5. Technical Report. Norrkøping, Sweden. URL: `http://www.hirlam.org/`.

Katz, R.W., Skaggs, R.H., 1981. On the use of autoregressive-moving average processes to model meteorological time series. Monthly Weather Review 109, 479–484.

Kristensen, N., Madsen, H., Jørgensen, S., 2004. Parameter estimation in stochastic grey-box models. Automatica 40, 225–237.

Lerch, S., Thorarinsdottir, T.L., 2013. Comparison of nonhomogeneous regression models for probabilistic wind speed forecasting. arXiv preprint arXiv:1305.2026 .

Ljung, L., 1999. System Identification. Wiley Online Library.

Madsen, H., 2008. Time series analysis. volume 72. CRC Press.

Messner, J.W., Zeileis, A., Broecker, J., Mayr, G.J., 2013. Probabilistic wind power forecasts with an inverse power curve transformation and censored regression. Wind Energy .

Mikosch, T., 1998. Elementary Stochastic Calculus: With Finance in View. volume 6. World Scientific.

Møller, J., Pinson, P., Madsen, H., 2013. Probabilistic Forecasts of Wind Power Generation by Stochastic Differential Equation Models. Technical Report. Technical University of Denmark. URL: `http://www.statistics.gov.hk/wsc/STS019-P5-S.pdf`.

Møller, J.K., Nielsen, H.A., Madsen, H., 2008. Time-adaptive quantile regression. Computational Statistics & Data Analysis 52, 1292–1303.

Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., Conzelmann, G., et al., 2009. Wind power forecasting: State-of-the-art 2009. Technical Report. Argonne National Laboratory (ANL).

Morales, J.M., Conejo, A.J., Madsen, H., Pinson, P., Zugno, M., 2014. Integrating Renewables in Electricity Markets – Operational Problems. volume 205 of *International Series in Operations Research & Management Science*. Springer, New York.

Nielsen, H.A., Nielsen, T.S., Madsen, H., Giebel, G., Badger, J., Landbergt, L., Sattler, K., Voulund, L., Tofting, J., 2006. From wind ensembles to probabilistic information about future wind power production–Results from an actual application, in: Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on, IEEE. pp. 1–8.

Øksendal, B., 2010. Stochastic Differential Equations: An Introduction with Applications. Universitext (1979), Springer. URL: http://books.google.dk/books?id=kXw9hB4EEpUC.

Pinson, P., 2012. Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions. Journal of the Royal Statistical Society: Series C (Applied Statistics) 61, 555–576.

Pinson, P., Girard, R., 2012. Evaluating the quality of scenarios of short-term wind power generation. Applied Energy 96, 12–20.

Pinson, P., Kariniotakis, G., 2010. Conditional prediction intervals of wind power generation. Power Systems, IEEE Transactions on 25, 1845–1856.

Pinson, P., Madsen, H., 2009. Ensemble-based probabilistic forecasting at horns rev. Wind energy 12, 137–155.

Pinson, P., Nielsen, H., Møller, J., Madsen, H., Kariniotakis, G., 2007. Nonparametric probabilistic forecasts of wind power: Required properties and evaluation. Wind Energy 10, 497–516.

Ramirez-Rosado, I.J., Fernandez-Jimenez, L.A., Monteiro, C., Sousa, J., Bessa, R., 2009. Comparison of two new short-term wind-power forecasting systems. Renewable Energy 34, 1848–1854.

Schefzik, R., Thorarinsdottir, T.L., Gneiting, T., et al., 2013. Uncertainty quantification in complex simulation models using ensemble copula coupling. Statistical Science 28, 616–640.

Taylor, J.W., McSharry, P.E., Buizza, R., 2009. Wind power density forecasting using ensemble predictions and time series models. Energy Conversion, IEEE Transactions on 24, 775–782.

The European Wind Energy Association, 2013. Ewea annual report 2012. Online. URL: http://www.ewea.org/publications/reports/ewea-annual-report-2012/.

Thorarinsdottir, T.L., Gneiting, T., 2010. Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression. Journal of the Royal Statistical Society: Series A (Statistics in Society) 173, 371–388.

Thorarinsdottir, T.L., Johnson, M.S., 2012. Probabilistic wind gust forecasting using nonhomogeneous Gaussian regression. Monthly Weather Review 140, 889–897.

Tol, R., 1997. Autoregressive conditional heteroscedasticity in daily wind speed measurements. Theoretical and applied climatology 56, 113–122.

Van Kampen, N.G., 1992. Stochastic Processes in Physics and Chemistry. volume 1. Access Online via Elsevier.

Welch, G., Bishop, G., 1995. An introduction to the Kalman filter. Online. URL: `http://clubs.ens-cachan.fr/krobot/old/data/positionnement/kalman.pdf`.

Zárate-Miñano, R., Anghel, M., Milano, F., 2013. Continuous wind speed models based on stochastic differential equations. Applied Energy 104, 42–49.

Zhou, Z., Botterud, A., Wang, J., Bessa, R., Keko, H., Sumaili, J., Miranda, V., 2013. Application of probabilistic wind power forecasting in electricity markets. Wind Energy 16, 321–338.

Paper E

Leveraging Stochastic Differential Equations for Probabilistic Forecasting of Wind Power using a Dynamic Power Curve

---

Submitted to *Wind Energy, 2014*.

WIND ENERGY
Wind Energ. 2015; **00**:1–11
DOI: 10.1002/we

RESEARCH ARTICLE

Leveraging Stochastic Differential Equations for Probabilistic Forecasting of Wind Power Using a Dynamic Power Curve

Emil B. Iversen, Juan M. Morales, Jan K. Møller, Pierre-Julien Trombe and Henrik Madsen

Technical University of Denmark, Asmussens Alle, building 303B, DK-2800 Lyngby, Denmark

ABSTRACT

Short-term (hours to days) probabilistic forecasts of wind power generation provide useful information about the associated uncertainty of these forecasts. Standard probabilistic forecasts are usually issued on a per-horizon-basis, meaning that they lack information about the development of the uncertainty over time or the inter-temporal correlation of forecast errors for different horizons. This information is very important for forecast end-users optimizing time-dependent variables or dealing with multi-period decision-making problems, such as the management and operation of power systems with a high penetration of renewable generation. This paper provides input to these problems by proposing a model based on stochastic differential equations that allows generating predictive densities as well as scenarios for wind power. We build upon a probabilistic model for wind speed and introduce a dynamic power curve. The model thus decomposes the dynamics of wind power prediction errors into wind speed forecast errors and errors related to the conversion from wind speed to wind power. We test the proposed model on an out-of-sample period of one year for a wind farm with a rated capacity of 21 MW. The model outperforms simple as well as advanced benchmarks on horizons ranging from 1 to 24 hours.

KEYWORDS
Wind Power; Dynamic Power Curve; Stochastic Differential Equations; Probabilistic Forecast

Received . . .

1. INTRODUCTION

Renewable energies have gained prominence in recent years as a sustainable solution to the world's growing energy demand. The largest share of renewable energy is produced by wind turbines in many countries (The European Wind Energy Association (2013)). The integration of wind power into the electricity grid is not without challenges. These are related to the variable and partly unpredictable behavior of the power generated by wind farms. To account for the stochastic nature of wind and to mitigate its potential adverse effects, accurate forecasts of future wind power generation are required. For the most efficient utilization of wind power, both in terms of grid stability and economic costs, forecasts that provide the full predictive distribution of the wind power generation are required (Morales et al. (2014); Pinson (2013)). Such forecasts are referred to as *probabilistic* in contrast to their deterministic counterpart, which consists of a single-value forecast, most typically the expected value or the most likely outcome.

1.1. Motivation

As motivation for the present research work, we illustrate the problem under consideration with a graphical example, see Figure 1. In Figure 1 the normalized wind power is plotted against the observed wind speed. These observations come from the Klim wind farm in Denmark. It is apparent that the relationship between observed wind speeds and generated power is not deterministic. Two sequences, each consisting of 15 data-points, are highlighted in red and blue, together with a local polynomial regression model of the relationship between wind speeds and generated power, shown in black, typically known as the power curve. The red sequence of data points indicate that the regressed curve consistently underestimates the power output. Conversely, the curve systematically overestimates the power output for the blue sequence of points. This reveals that the relationship between observed wind speed and observed power output exhibits some memory and may change over time. This may be attributed to a variety of factors such as the turbine blades being dirty, different local

Prepared using weauth.cls [Version: 2010/06/17 v1.00]

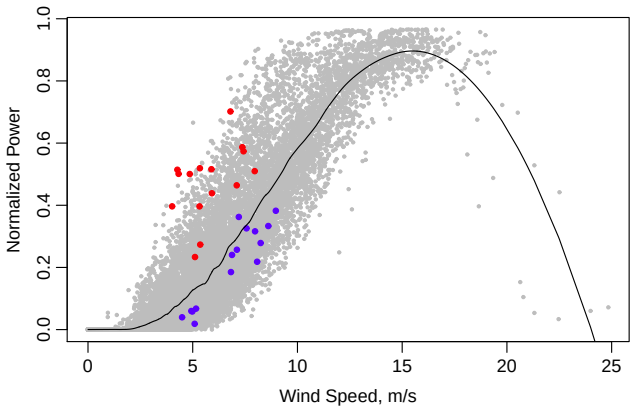

Figure 1. *Normalized power plotted against wind speeds in gray both in hourly averages. The black line indicates a local polynomial regression model of the relationship between wind speeds and normalized power. The red points represent a sequence of 15 hourly observations. The blue dots are a sequence of 15 hourly observations 9 months later.*

turbulence characteristics (tree foliage, wind shadowing) or the wind direction. These features have also been pointed out by other authors for other data sets (Jeon and Taylor (2012), Wächter et al. (2011)). The key takeaway, however, remains the same: If the power curve is to be used for forecasting, it should be dynamic and adaptive to accommodate the changes in the relationship between wind speed and the generated power.

There exists a wealth of approaches for generating wind power forecasts, albeit mostly centered on point predictions. Thorough reviews of the most relevant works in this field are given in Costa et al. (2008); Monteiro et al. (2009) and Foley et al. (2012).

Several studies document and emphasize the importance of probabilistic forecasts for wind power integration. Among such examples are Pinson et al. (2007), which illustrates the value of such forecasts for energy trading. In Bludszuweit and Domínguez-Navarro (2011) the usefulness of probabilistic forecasts for operating energy storage is highlighted. Another example is Matos and Bessa (2011), where determining the grid operating reserve is in focus. In response to this need for describing the uncertainty in the wind power forecast, probabilistic forecasts have received increased attention in recent years with the application of methods such as quantile regression (Bremnes (2004); Møller et al. (2008)) and the use of ensemble forecasts (Pinson and Madsen (2009)) and Box-Jenkins models (Trombe et al. (2012)). Furthermore, forecasters are also focusing on generating scenarios and constructing models that capture the interdependence in time of forecast errors (Pinson et al. (2009), Bessa et al. (2012), Møller et al. (2013)).

Power curves for individual wind turbines are usually supplied by the wind turbine manufacturer. This power curve is typically deterministic. For wind farms, the aggregate power curve depends on many other factors, such as the topography, wind shading and wind direction. A stochastic power curve has been modeled in Jeon and Taylor (2012), where the time variation of the curve due to some of these factors is captured by an adaptive estimation procedure. A Markovian power curve is used in Anahua et al. (2008) to simulate wind power fluctuations. In Gottschall and Peinke (2007) a stochastic power curve is also shown to yield a more accurate model of the wind power generation.

The main contribution of this paper is the development of a dynamic power curve model that allows us to translate a probabilistic forecast of wind speed into a probabilistic forecast of wind power. For this purpose, the proposed power curve model is both probabilistic and dynamic, so as to reflect the changing characteristics of the wind farm. This is achieved by setting up our model in the form of stochastic differential equations.

The remainder of this paper is organized as follows: Section 2 gives a very brief introduction to stochastic differential equations and how to estimate their characteristic parameters. In Section 3 we present two stochastic differential equation models for wind power, which differ in whether wind speed observations are used or not. The following section, Section 4, deals with the performance of the generated forecasts. Section 5 concludes the paper and outlines future research.

Wind Energ. 2015; **00**:1–11 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

2. STOCHASTIC DIFFERENTIAL EQUATIONS

Stochastic differential equations (SDEs) extend ordinary differential equations (ODEs) by including one or more stochastic terms. The solution to an SDE is, therefore, a stochastic process. SDE models have been employed in a variety of fields to describe systems with a large stochastic component. This type of models has been widely used in mathematical finance (Björk (2009); Mikosch (1998)) and physics (Van Kampen (1992); Adomian (1988)). More recently, such models have also been used for modeling and forecasting solar irradiance (Iversen et al. (2014a)) and wind speed (Iversen et al. (2014b)). In this section we provide a very succinct description of SDEs and refer the reader to Øksendal (2010) or Jazwinski (2007) for general discussions on the topic and to Kristensen et al. (2004) for more details on the specific approach followed in this paper.

The standard notation for SDEs is similar to that of ODEs and takes the following form.

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t, \tag{1}$$

where W_t is a Wiener process, and $f(\cdot)$ and $g(\cdot)$ are known as the drift and diffusion terms, respectively. Owing to the fact that the derivative of W_t, dW_t, does not exist, this equation should be interpreted as a convenient way of writing the following integral equation:

$$X_t = X_0 + \int_0^t f(X_s, s)ds + \int_0^t g(X_s, s,)dW_s, \tag{2}$$

where the Itō interpretation of the second integral should be used, see Øksendal (2010).

Solving the SDE yields a stochastic process, which characterizes the uncertainty in the process dynamics for every future time. The probability density function of the stochastic process at a certain time in the future can be obtained by solving a partial differential equation referred to as the Fokker-Plank or forward Kolmogorov equation, see Björk (2009) for details. In general, this solution cannot be found analytically; however, a significant number of numerical solution approaches are currently available, see e.g. Jazwinski (2007).

A convenient result about the generality of SDEs is the Lévy-Itō decomposition, see Björk (2009), which essentially states that all stochastic processes with continuous trajectories can be written as special cases of SDEs. Indeed, many ordinary time-series models can be seen as discrete time interpretations of SDEs.

Though continuous in nature, in practice SDEs can only be observed or measured at discrete instances in time. For this reason, an *observation* or *measurement* equation, $h(\cdot)$, is introduced into the SDE system, which then becomes:

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t \tag{3}$$

$$Y_k = h(X_{t_k}, t_k, e_k). \tag{4}$$

Notice that we have now defined the observation Y_k of the process at time t_k. This measurement equation also allows us to specify an observation error term, e_k, in our model.

The procedure that we use to estimate the SDE parameters hinges on a specific form of the observation equation, $h(\cdot)$, and the diffusion term, $g(\cdot)$. In particular, we require $g(X_t, t) = g(t)$ and $h(X_{t_k}, t_k, e_k) = h(X_{t_k}, t_k) + e_k$, where $e_k \sim \mathcal{N}(0, \sigma^2)$. In principle, these conditions limit our modeling framework significantly. However, they can be relaxed to a large extent. Indeed, the requirement that $g(\cdot)$ does not depend on X_t can be overcome by making use of Itō calculus to transform the original process with state-dependent diffusion to an equivalent one with *non*-state-dependent diffusion, see e.g.Møller et al. (2008); Iversen et al. (2014a). The second requirement, which implies that the observation error is additive and normally distributed, can be relieved by transforming the data, see e.g. Box and Cox (1964).

The method we use for parameter estimation has been outlined in several previous papers and therefore, we will not describe it here. Instead, we refer the reader to Jazwinski (2007) for a general discussion on parameter estimation for SDEs and to Møller and Madsen (2010), Iversen et al. (2014b) and Juhl et al. (2013) for practical guidance on its application, which also includes an open source implementation. Note that the estimation procedure, as implemented in the R package, entails maximizing the likelihood function over the residuals computed one step ahead. We follow suit in this paper. Nonetheless, estimating SDE models using a multi-horizon approach is also possible, but at a high computational cost, see Møller et al. (2013), for instance.

3. MODELS FOR WIND POWER

In this section we propose two models for wind power, one that assumes that both wind power and wind speed are measured and one where we only observe wind power. There are several reasons that motivate these two approaches: first of all, not all wind farms have access to wind speed data. Second, the difference in performance of the two models may highlight the benefit of having wind speed measurements available. Third, not observing wind speed allows us to investigate the feasibility of constructing a dynamic power curve model in such a case.

3.1. Model with Wind Speed Observations

We start out with the following probabilistic model for wind speed,

$$dX_t = \left(\left(1 - e^{-X_t}\right)(\rho_x \dot{p}_t + R_t) + \theta_x(p_t\mu_x - X_t)\right)dt + \sigma_x X_t^{0.5} dW_t \quad (5)$$

$$dR_t = -\theta_r R_t dt + \sigma_r dW_t \quad (6)$$

$$Y_k = X_{t_k} + \epsilon_k, \quad (7)$$

which is a simplified version of the model developed in Iversen et al. (2014b). X_t, R_t, Y_k and ϵ_k are stochastic variables, with X_t being the actual wind speed, Y_t being the observed wind speed, R_t can be interpreted as a dynamic correction of the Numerical Weather Prediction (NWP) p_t, and $\epsilon_k \sim \mathcal{N}(0, \sigma_\epsilon)$ is an observation error. The first order differentiation in the NWP of wind speed is introduced as $\dot{p}_t$. The term $\left(1 - e^{-X_t}\right)$ is included to keep the process in the proper domain, $\mathbb{R}^+$. ρ_x governs the emphasis put on $\dot{p}_t$. θ_x and θ_r govern the speed of reversion to $\mu_x p_t$ and zero for X_t and R_t, respectively. μ_x corrects the NWP from systematically over or under shooting the observed values of wind speed. σ_x, σ_r and σ_ϵ govern the size of the diffusions of X and R and the size of the observation error, respectively. The model described here succeeds in describing the dynamics of the wind speeds at a specific location. The diffusion term, $\sigma_x X_t^{0.5}$ drops to zero when X_t is close to zero. In such a case, the drift term dominates and pushes the process away from zero with the result that X_t always takes on values within the domain $\mathbb{R}^+$.

Since we presume that a large part of the variation in the wind power output can be explained by the underlying wind speed process, we aim at capturing the speed-to-power relationship by introducing a dynamic power curve in the SDE system (5)–(7). This results in the following probabilistic model for wind power.

$$dX_t = \left(\left(1 - e^{-X_t}\right)(\rho_x \dot{p}_t + R_t) + \theta_x(p_t\mu_x - X_t)\right)dt + \sigma_x X_t^{0.5} dW_{x,t} \quad (8)$$

$$dR_t = -\theta_r R_t dt + \sigma_r dW_{r,t} \quad (9)$$

$$dQ_t = (S_t - \theta_q Q_t)dt + \sigma_q dW_{q,t} \quad (10)$$

$$dS_t = -\theta_s S_t dt + \sigma_s dW_{s,t} \quad (11)$$

$$Y_{1,k} = X_{t_k} + \epsilon_{1,k} \quad (12)$$

$$Y_{2,k} = (0.5 + 0.5\tanh(5(X_{t_k} - \gamma_1)))\,(0.5 - 0.5\tanh(\gamma_2(X_{t_k} - \gamma_3))) \quad (13)$$

$$\frac{\zeta_3}{1 + e^{-\zeta_1(X_{t_k} - \zeta_2 + Q_{t_k})}} + \epsilon_{2,k}, \quad (14)$$

where equations (8),(9) and (12) correspond to our previous wind speed model. Notice that we have included a new observation, $Y_{2,k}$, which is the normalized power. The functional relationship between the wind speed and the normalized power is partly determined by the parameters γ_1, γ_2 and γ_3, which model the cut-in and cut-out wind speeds and how fast the wind turbines cut-out, and partly by ζ_1, ζ_2 and ζ_3, which govern the slope of the power curve in the domain where the wind turbines produce power. We also introduce the stochastic variable Q_t to capture the fact that the speed-to-power efficiency of the wind farm changes over time in a stochastic manner. This feature of the model is based on observations from the data, as shown in Figure 1, which may be caused by factors such as dirty blades, how laminar the wind flow is, turbine failures or turbine maintenance, all of which strongly affect the rate at which wind is converted into power. Thus, we add dynamics to the power curve to account for this shift over time through Q_t. S_t is a stochastic variable that helps model the dynamics of Q_t. Along with S_t and Q_t we introduce the parameters $\theta_q, \theta_s, \sigma_q$ and σ_s that describe the evolution of these stochastic variables.

3.2. Model Without Wind Speed Observations

An obvious variant of the former wind power model results from considering that wind speed measurements are not directly available (e.g., through an anemometer) and thus, only the power output of the wind farm is observed. This model would be useful if we were to apply the model on a wind farm where we do not have access to wind speed observations. This could, in principle, be done by simply removing equation (12) from the above model. However, if we proceeded in this way, we would end up with an unidentifiable model, the parameters of which could not be estimated and for which the underlying states could not be filtered. The idea here is that we can decompose model (8)–(14) into a system with fast dynamics (the wind speed process) and a system with slow dynamics (the conversion from speed to power). However, if we do not observe the wind speed, this decomposition turns out to be infeasible. Ljung (1999) addresses the topic of identifiability for stochastic models. Consequently, in the absence of wind speed measurements, the SDE model (8)–(14)

Wind Energ. 2015; **00**:1–11 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

has to be simplified substantially, resulting in the following model:

$$dX_t = \left(\left(1 - e^{-X_t}\right)(\rho_x \hat{p}_t + R) + \theta_x(p_t\mu_x - X_t)\right)dt + \sigma_x X_t^{0.5} dW_t \quad (15)$$

$$dR_t = -\theta_r R_t dt + \sigma_r dW_t \quad (16)$$

$$Y_{2,k} = (0.5 - 0.5\tanh(\gamma_2(X_{t_k} - \gamma_3)))\frac{\zeta_3}{1 + e^{-\zeta_1(X_{t_k} - \zeta_2)}} + \epsilon_{2,k}. \quad (17)$$

In the model specified by equations (15)–(17), the power curve (17) is not dynamic as it remains unchanged over time. In this case, X_t can no longer be interpreted as the wind speed per se, but instead as a state variable that we may call the *effective wind speed*, that is, the wind speed that would yield the measured power output given the deterministic power curve defined in equation (17).

4. MODEL EVALUATION

In this section, we focus on validating the proposed model (8)–(14) while also considering the difference between this model and model (15)–(17).

The data used in this study originates from the Klim Fjordholme wind farm and consists of power measurements and predicted wind speeds based on a numerical weather prediction (NWP) model from the Danish Meteorological Institute (Källén (1996)). The data is sampled hourly. The power data from the Klim wind farm has been used in previous studies, see for instance Pinson (2013). The NWP model provides a wind speed forecast up to 48-hours ahead and is updated every 6 hours. The data covers three years, from 01/01-1999 to 31/12-2001 and is divided into two periods: a training set spanning two years, which is used for estimation, and a test set covering the remaining one year to evaluate the performance of the proposed model.

The first step in the validation task is to inspect the parameter estimates of the model, which are easily interpretable and, as such, provide a quick check of how (un)reasonable our model is at a first glance. The parameter values are shown in Table I.

$\hat{\gamma}_1$	2.648	$\hat{\sigma}_q$	1.093	$\hat{\rho}_x$	0.3087
$\hat{\gamma}_2$	0.657	$\hat{\theta}_r$	1.358	$\hat{\mu}_x$	0.7120
$\hat{\gamma}_3$	19.63	$\hat{\sigma}_r$	1.476	$\hat{\theta}_x$	0.2317
$\hat{\zeta}_1$	0.612	$\hat{\theta}_s$	0.0629	$\hat{\sigma}_x$	0.0884
$\hat{\zeta}_2$	9.312	$\hat{\sigma}_s$	0.7195	$\hat{\sigma}^2_{\epsilon_1}$	0.1324
$\hat{\zeta}_3$	0.952	$\hat{\theta}_q$	1.788	$\hat{\sigma}^2_{\epsilon_2}$	0.0001259

Table I. *Parameter estimates for model (8)–(14)*

The shape of the stationary power curve, which is depicted in Figure 2 along with the observations, is determined by the parameters $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\zeta}_1, \hat{\zeta}_2$ and $\hat{\zeta}_3$. Notice that the cut-out wind speed, which is given by γ_3 in the model, is estimated to be 19.63 m/s, that is, close to the rated cut-out wind speed for the Vestas V44 wind turbines installed at Klim fjordholme, which is 20 m/s. It is important to stress that the power curve that is modeled here corresponds to that of the wind farm and not to the power curve for the individual wind turbine. We have normalized the wind farm power output with its rated capacity of 21 MW. However, the actual nominal maximal production seems closer to $\hat{\zeta}_3 \cdot 21 = 20.00$ MW, which may be due to turbine aging. Notice also that $\hat{\mu}_x$ is somewhat close to one indicating that the predicted wind speed for longer horizons should tend to the numerical weather prediction, albeit with some bias.

To give an indication of the SDE model performance, we compare it against some simple and some more advanced benchmarks. We focus on standard time-series benchmarks ranging from a persistence model to an ARX-GARCH. The persistence model of order j is given by

$$Y_k = Y_{k-j} + \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}(0, \sigma^2). \quad (18)$$

While a persistence model does not require any specific associated distribution of the innovation term, we assume that it is normally distributed to fit into the classical AR model setup and to compute the continuous ranked probability score for the persistence model.

We specify an auto-regressive (AR) model of order q as

$$Y_k = \psi_0 + \sum_{i=1}^{q} \psi_i Y_{k-i} + \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}(0, \sigma^2). \quad (19)$$

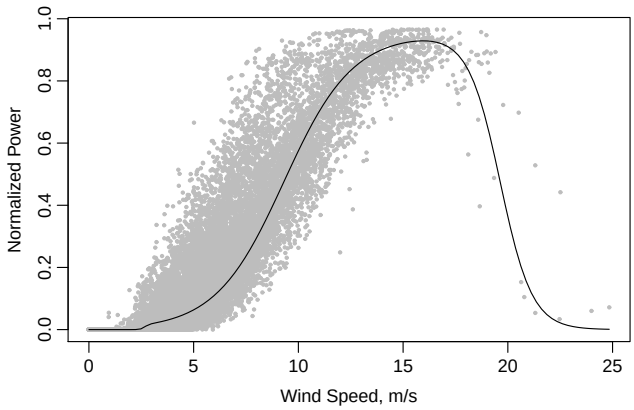

Figure 2. *The power curve obtained when the states Q_t and S_t are at their long term stationary levels.*

An auto-regressive with external input (ARX) model takes the form

$$Y_k = \psi_0 + \sum_{i=1}^{q} \psi_i Y_{k-i} + \phi p_{t_k} + \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}(0, \sigma^2), \tag{20}$$

where p_{t_k} is the numerical weather prediction at time t_k. We also consider an ARX model with truncated normal distributed innovation (ARX-TN), where ϵ_k follows a truncated normal, such that the process is confined to $[0, 1]$.

Lastly, we consider an auto-regressive model with external input and a generalized auto-regressive conditional heteroskedastic variance term (ARX-GARCH) which takes the following form:

$$Y_k = \psi_0 + \sum_{i=1}^{q} \psi_i Y_{k-i} + \phi p_{t_k} + \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}(0, \sigma_k^2) \tag{21}$$

$$\sigma_k^2 = \alpha_0 + \sum_{i=1}^{\tilde{q}} \alpha_i \sigma_{k-i}^2 + \sum_{i=j}^{\tilde{p}} \beta_j \tilde{\epsilon}_{k-j}^2 + \tilde{\epsilon}_k, \qquad \tilde{\epsilon}_k \sim \mathcal{N}(0, \tilde{\sigma}^2), \tag{22}$$

where the normally distributed innovation, $\epsilon_k \sim \mathcal{N}(0, \sigma_k^2)$ is truncated such that the process is confined to $[0, 1]$ for the variant of this benchmark with truncated innovation (ARX - GARCH - TN).

Models with the same general structure as the benchmarks above have been used to forecast wind power production in several scientific publications such as Duran et al. (2007), Jeon and Taylor (2012), Taylor et al. (2009), Lau and McSharry (2010), Trombe et al. (2012).

		Test Set		
Models	Parameters	MAE	RMSE	CRPS
Climatology	-	0.2208	0.2693	0.1417
Persistence	1	0.0509	0.0835	0.0428
AR	4	0.0527	0.0820	0.0417
ARX	5	0.0510	0.0795	0.0406
ARX - TN	7	0.0648	0.0848	0.0444
ARX - GARCH	9	0.0505	0.0797	0.0382
ARX - GARCH - TN	11	0.0575	0.0823	0.0401
Model (8)–(14)	19	**0.0471**	**0.0773**	**0.0327**
Model (15)–(17)	12	0.0553	0.08988	0.0399

Table II. *The MAE, RMSE and CRPS scores for benchmarks as well as for the proposed model.*

Wind Energ. 2015; **00**:1–11 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

In Table II, the performance for the benchmarks and the proposed SDE model are shown for 1-step (i.e. 1 hour) ahead forecasts. We evaluate in terms of various scores: the Mean Absolute Error (MAE) which is computed as

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \hat{Y}_i\right|, \tag{23}$$

and the Root Mean Squared Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2}, \tag{24}$$

where N is the number of observations, Y_i is the i'th wind power observation and $\hat{Y}_i$ is the corresponding forecast power. Also, we use the continuous ranked probability score (CRPS) as defined in Gneiting and Raftery (2007) to rank and compare the models in a probabilistic sense. We see here that the SDE model (8)–(14) outperforms simple as well as stronger benchmarks. Comparing model (8)–(14) and model (15)–(17) we note that there is a clear benefit from having wind speed observations along with a dynamic power curve, this is both in terms of the point forecast, as measured by the MAE and the RMSE, as well as in terms of the predictive density, as measured by the CRPS.

Figure 3 shows the 1-hour and 24-hour ahead predictive densities of wind power given by the proposed SDE model. Three observations are in order. First, note that the predictive densities are more spread out for the 24-hour ahead forecasts, revealing a higher degree of uncertainty for longer prediction horizons, as expected. Second, observe that the density is also more spread out when the predicted normalized power is around 0.5. This follows from the fact that this is the power value around which the power curve has the steepest slope and therefore, a certain variation in wind speed will, all other things equal, yield a larger variation in the power produced. Third, notice that the predictive density seems to become quite sharp around zero predicted power. This can be explained by means of Figure 2, where we can see that the power curve is essentially flat and equal to zero for low wind speeds, thus yielding zero power and very little variation for low wind speed values. This is a clear advantage over Box-Jenkins-type models, as the (8)–(14) model can associate a NWP of small wind speeds with very little variability in the prediction of generated power.

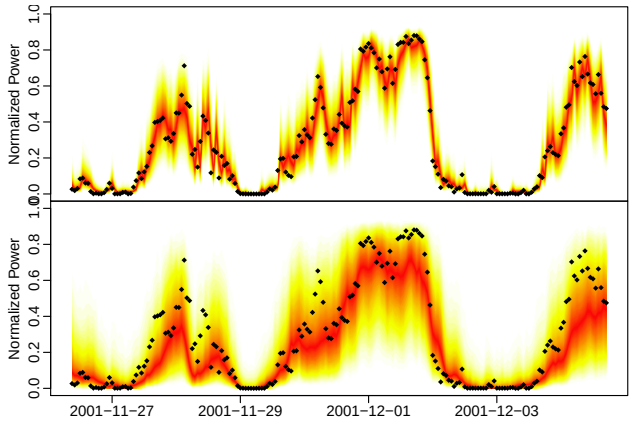

Figure 3. *1-hour (top) and 24-hour (bottom) ahead densities predicted by model (8)–(14), with warmer colors indicating a higher probability of seeing this realization. The densities are approximated by Monte Carlo simulations.*

As hinted in Figure 3, SDEs automatically provide a framework for providing multi-horizon forecast densities. In this line, Figure 4 displays the predictive densities for 1-24 hours ahead.

It is important to note, however, that, even though the proposed SDE model readily provides predictive densities for horizons beyond the first hour, it is not designed to be used in this way as it is estimated based one-step ahead residuals. Notwithstanding this, we evaluate next the performance of the proposed SDE model beyond 1-hour horizons and compare

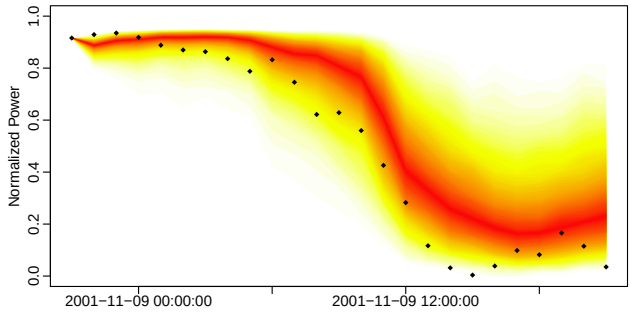

Figure 4. *Multi-horizon predictive densities for normalized wind power generation from model (8)–(14). Warmer colors indicate a higher probability of seeing this realization. The density is approximated by Monte Carlo simulations.*

Models	CRPS for different horizons				Energy Score
	1 hour	4 hours	12 hours	24 hours	
ARX - GARCH	0.0382	0.0704	0.0787	**0.0789**	1.180
ARX - GARCH - TN - iterative	0.0401	0.0783	0.1043	0.1225	1.945
Model (8)–(14)	**0.0327**	**0.0641**	**0.0779**	0.0836	**0.739**
Model (15)–(17)	0.0399	0.0685	0.0784	0.0793	0.745

Table III. *The CRPS and energy score for SDE models (8)–(14) and (15)–(17) and for the ARX-GARCH benchmarks. The ARX-GARCH benchmark is fitted specifically for each forecast horizon. The iterative ARX-GARCH-TN model is fitted to 1-hour ahead data and then run iteratively until the desired horizon is reached.*

it with the best performing benchmarks on these same horizons. The results of this comparison are collated in Table III. We compare in terms of the CRPS and the energy score, both defined in Gneiting and Raftery (2007) and we consider two benchmarks. First, an ARX-GARCH model, which is specified as previously, but now fitted specifically to each prediction horizon. On the one hand, this confers an advantage on this benchmark model with respect to the proposed one, because, as we have just mentioned, the SDE model is only fitted for 1-hour ahead data and forecasts for longer horizons are obtained by extrapolation. On the other hand, the so-fitted ARX-GARCH model fails to capture the time structure of forecast errors. This brings us to the second benchmark, the ARX-GARCH-TN-iterative, where we run the ARX-GARCH-TN in an iterative fashion to obtain forecasts for the desired horizons. This benchmark is able to provide both trajectories and multi-horizon forecasts. We choose the truncated version of the ARX-GARCH model, as the standard ARX-GARCH model becomes unstable for a large number of iterations.

From Table III we see that model (8)–(14) outperforms the two benchmarks for all lead times shorter than 24 hours in terms of the CRPS. However, the ARX-GARCH benchmark performs the best on the 24-hour horizon. We insist, though, that the ARX-GARCH model is specifically designed to provide power forecasts on the desired horizon, while failing to produce multi-horizon forecasts and time-path trajectories with the proper time structure. The proposed model (8)–(14) also outperforms the two benchmarks in terms of the energy score. Further insight can be gained from Table III by comparing model (8)–(14) and model (15)–(17). The former exploits wind speed observations, while the latter does not. Notice that including wind speed observations improves the forecast significantly for short horizons, but this is not the case for longer horizons. This is so because information on recently past wind speed proves to be useful to identify and predict the range of the power curve in which the wind farm is operating. In practice this allows a more advantageous use of the numerical weather prediction for predicting wind speeds. Naturally, the predictive power of wind speed measurements is first diluted and finally vanishes with longer prediction horizons.

Other approaches exist for producing multi-horizon forecasts of wind power production. One method that is particularly popular is to fit a marginal predictive density for each specific lead time and then to combine them into a cumulative distribution using copulas (Schefzik et al. (2013); Bessa et al. (2012)). Copula methods are, however, computationally demanding and require estimation of a large number of parameters. Another approach is to simultaneously fit a SDE model to different horizons, thus addressing the issue of the large number of parameters to be estimated (Møller et al. (2013)).

Wind Energ. 2015; **00**:1–11 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

However, this method is still computationally burdensome and can only handle a limited amount of data. Hence, we claim that our SDE approach, although relying on a one-step-ahead estimation procedure, can be considered a reasonable method for multi-horizon forecasting; it requires relatively few parameters and is comparable in computation time to standard time-series models.

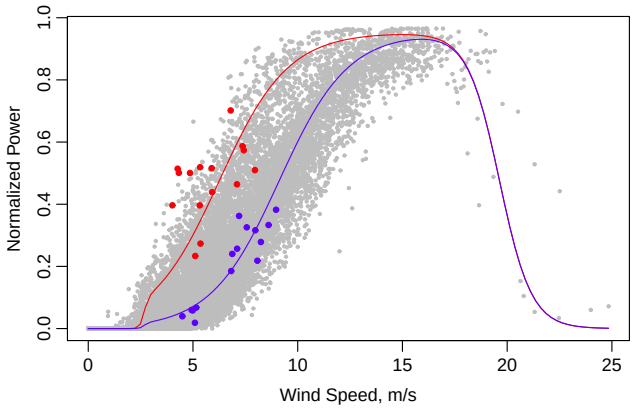

Figure 5. *Normalized power plotted against wind speeds in gray both in hourly averages, similar to Figure 1. The blue and red lines are the average power curves from model (8)–(14) for the times when the blue and the red observations were recorded.*

To conclude this section, we plot again (see Figure 5) the two sets of 15 data points with 9 months between them that we showed in Figure 1 in the introduction. This time, however, instead of the power curve fitted with the local regression in Figure 1, we depict the adaptive power curve that is specific for each time period. The resulting power curves are shown in Figure 5. As it turns out, these power curves seem to much better capture the time-varying relationship between wind speed and wind power.

5. CONCLUDING REMARKS

Probabilistic forecasts of wind power generation provide useful input information to a variety of decision-making problems concerning management and trading of wind-generated electricity. Standard forecasts do not take into account the development of the uncertainty over time, making them of little use for a large class of operational tasks which involve time-dependent decisions. Such problems are related to the large-scale integration of wind power in electrical energy systems, where the dynamics of the uncertainty in the wind generation is of critical importance.

In this paper we suggest a new framework for modeling, simulating and forecasting wind power generation for multiple horizons. The proposed framework provides wind power forecasts by combining a dynamic power curve with a stochastic model for wind speed based on stochastic differential equations. This results in an automatically bounded wind power process with time-varying uncertainty. The proposed model outperforms simple as well as complex benchmarks on an out-of-sample period of one year on horizons ranging from 1-24 hours.

Using stochastic differential equations for forecasting lends itself well to several extensions. One such an extension is to consider spatio-temporal forecasting for capturing the interdependence between different sites in space and time. Another extension is to consider forecasting multiple outputs such as wind power generation, solar power generation and perhaps power load in a single complete model.

ACKNOWLEDGMENTS

DSF (Det Strategiske Forskningsråd) is to be acknowledged for partly funding the work of Emil B. Iversen, Juan M. Morales and Henrik Madsen through the Ensymora project (no. 10-093904/DSF). Furthermore, Juan M. Morales and

Henrik Madsen are partly funded by the Research Center CITIES (no. 1035-00027B), which is also supported by DSF (Det Strategiske Forskningsråd).

REFERENCES

Adomian, G., 1988. Nonlinear Stochastic Systems Theory and Application to Physics. Vol. 46. Springer.

Anahua, E., Barth, S., Peinke, J., 2008. Markovian power curves for wind turbines. Wind Energy 11 (3), 219–232.

Bessa, R. J., Miranda, V., Botterud, A., Zhou, Z., Wang, J., 2012. Time-adaptive quantile-copula for wind power probabilistic forecasting. Renewable Energy 40 (1), 29–39.

Björk, T., 2009. Arbitrage Theory in Continuous Time. Oxford Finance Series. OUP Oxford.
URL http://books.google.dk/books?id=N0ImZs3HpwIC

Bludszuweit, H., Domínguez-Navarro, J. A., 2011. A probabilistic method for energy storage sizing based on wind power forecast uncertainty. Power Systems, IEEE Transactions on 26 (3), 1651–1658.

Box, G. E., Cox, D. R., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological), 211–252.

Bremnes, J. B., 2004. Probabilistic wind power forecasts using local quantile regression. Wind Energy 7 (1), 47–54.

Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H., Feitosa, E., 2008. A review on the young history of the wind power short-term prediction. Renewable and Sustainable Energy Reviews 12 (6), 1725–1744.

Duran, M. J., Cros, D., Riquelme, J., 2007. Short-term wind power forecast based on ARX models. Journal of Energy Engineering 133 (3), 172–180.

Foley, A. M., Leahy, P. G., Marvuglia, A., McKeogh, E. J., 2012. Current methods and advances in forecasting of wind power generation. Renewable Energy 37 (1), 1–8.

Gneiting, T., Raftery, A., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 (477), 359–378.

Gottschall, J., Peinke, J., 2007. Stochastic modelling of a wind turbine's power output with special respect to turbulent dynamics. In: Journal of Physics: Conference Series. Vol. 75. IOP Publishing, p. 012045.

Iversen, E. B., Morales, J. M., Møller, J. K., Madsen, H., 2014a. Probabilistic Forecasts of Solar Irradiance by Stochastic Differential Equations. Environmetrics 25 (1), 152–164.

Iversen, E. B., Morales, J. M., Møller, J. K., Madsen, H., 2014b. Short-term Probabilistic Forecasting of Wind Speed Using Stochastic Differential Equations. Accepted for publication in International Journal of Forecasting(in press).

Jazwinski, A. H., 2007. Stochastic Processes and Filtering Theory. Courier Dover Publications.

Jeon, J., Taylor, J. W., 2012. Using conditional kernel density estimation for wind power density forecasting. Journal of the American Statistical Association 107 (497), 66–79.

Juhl, R., Kristensen, N. R., Bacher, P., Mø ller, J. K., Madsen, H., 2013. CTSM-R user guide. Technical University of Denmark.

Källén, E., 1996. Hirlam documentation manual; 1996. system 2.5. Tech. rep., Norrkøping, Sweden.
URL http://www.hirlam.org/

Kristensen, N., Madsen, H., Jørgensen, S., 2004. Parameter estimation in stochastic grey-box models. Automatica 40 (2), 225–237.

Lau, A., McSharry, P., 2010. Approaches for multi-step density forecasts with application to aggregated wind power. The Annals of Applied Statistics 4 (3), 1311–1341.

Ljung, L., 1999. System Identification. Wiley Online Library.

Matos, M. A., Bessa, R. J., 2011. Setting the operating reserve using probabilistic wind power forecasts. Power Systems, IEEE Transactions on 26 (2), 594–603.

Mikosch, T., 1998. Elementary Stochastic Calculus: With Finance in View. Vol. 6. World Scientific.

Møller, J., Madsen, H., 2010. From state dependent diffusion to constant diffusion in stochastic differential equations by the Lamperti transform. Tech. rep., Technical University of Denmark.
URL http://orbit.dtu.dk/fedora/objects/orbit:82985/datastreams/file_5175845/content

Møller, J., Pinson, P., Madsen, H., 2013. Probabilistic forecasts of wind power generation by stochastic differential equation models. Tech. rep., Technical University of Denmark.
URL http://www.statistics.gov.hk/wsc/STS019-P5-S.pdf

Møller, J. K., Nielsen, H. A., Madsen, H., 2008. Time-adaptive quantile regression. Computational Statistics & Data Analysis 52 (3), 1292–1303.

Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., Conzelmann, G., et al., 2009. Wind power forecasting: State-of-the-art 2009. Tech. rep., Argonne National Laboratory (ANL).

Morales, J. M., Conejo, A. J., Madsen, H., Pinson, P., Zugno, M., 2014. Integrating Renewables in Electricity Markets – Operational Problems. Vol. 205 of International Series in Operations Research & Management Science. Springer, New York.

Øksendal, B., 2010. Stochastic Differential Equations: An Introduction with Applications. Universitext (1979). Springer.
URL http://books.google.dk/books?id=kXw9hB4EEpUC

Pinson, P., 2013. Wind energy: Forecasting challenges for its operational management. Statistical Science 28 (4), 564–585.

Pinson, P., Chevallier, C., Kariniotakis, G. N., 2007. Trading wind generation from short-term probabilistic forecasts of wind power. Power Systems, IEEE Transactions on 22 (3), 1148–1156.

Pinson, P., Madsen, H., 2009. Ensemble-based probabilistic forecasting at horns rev. Wind Energy 12 (2), 137–155.

Pinson, P., Madsen, H., Nielsen, H. A., Papaefthymiou, G., Klöckl, B., 2009. From probabilistic forecasts to statistical scenarios of short-term wind power production. Wind energy 12 (1), 51–62.

Schefzik, R., Thorarinsdottir, T. L., Gneiting, T., et al., 2013. Uncertainty quantification in complex simulation models using ensemble copula coupling. Statistical Science 28 (4), 616–640.

Taylor, J. W., McSharry, P. E., Buizza, R., 2009. Wind power density forecasting using ensemble predictions and time series models. Energy Conversion, IEEE Transactions on 24 (3), 775–782.

The European Wind Energy Association, June 2013. EWEA annual report 2012. Online.
URL http://www.ewea.org/publications/reports/ewea-annual-report-2012/

Trombe, P.-J., Pinson, P., Madsen, H., 2012. A general probabilistic forecasting framework for offshore wind power fluctuations. Energies 5 (3), 621–657.

Van Kampen, N. G., 1992. Stochastic Processes in Physics and Chemistry. Vol. 1. Access Online via Elsevier.

Wächter, M., Milan, P., Mücke, T., Peinke, J., 2011. Power performance of wind energy converters characterized as stochastic process: applications of the langevin power curve. Wind Energy 14 (6), 711–717.

Paper F

Spatio-Temporal Forecasting by Coupled Stochastic Differential Equations: Applications to Solar Power

Submitted to *Journal of American Statistical Association, 2015.*

Spatio-Temporal Forecasting by Coupled Stochastic Differential Equations: Applications to Solar Power

Emil B. Iversen[a], Rune Juhl[a], Juan M. Morales[a], Jan K. Møller[a], Jan Kleissl[b], Henrik Madsen[a]

[a] *Technical University of Denmark, Asmussens Alle, building 322, DK-2800 Lyngby, Denmark*
[b] *University of California, San Diego, 9500 Gilman Drive MC 0411, La Jolla, CA 92093, USA*

Abstract

Spatio-temporal problems exist in many areas of knowledge and disciplines ranging from biology to engineering and physics. However, solution strategies based on classical statistical techniques often fall short due to the large number of parameters that are to be estimated and the huge amount of data that need to be handled. In this paper we apply known techniques in a novel way to provide a framework for spatio-temporal modeling which is both computationally efficient and has a low dimensional parameter space. We present a micro-to-macro approach whereby the local dynamics are first modeled and subsequently combined to capture the global system behavior. The proposed methodology relies on coupled stochastic differential equations and is applied to produce spatio-temporal forecasts for a solar power plant for very short horizons, which essentially implies tracking clouds moving across the field of solar power inverters. We outperform simple and complex benchmarks while providing forecasts for 70 spatial dimensions and 24 lead times (i.e., for a total number of random variables equal to 1680). The resulting model can provide all sorts of forecast products, ranging from point forecasts and co-variances to predictive densities, multi-horizon forecasts, and space-time trajectories.

Keywords: Spatio-temporal modeling, Stochastic partial differential equations, Solar power forecasting, Nowcasting, Cloud tracking

1. Introduction

Recent years have seen a massive increase in the collection of data. The challenges related to the treatment of large datasets have, in turn, been the subject of intense research. Among these challenges are the development of predictive methods for spatio-temporal systems. In this article we are concerned with modeling the spatio-temporal behavior in order to better predict the aggregate system dynamics. The need for high dimensional predictions arises in many fields. In meteorology and hydrology such methods allow for precipitation nowcasting to predict flooding and to predict extreme wind speeds ([36], [30], [38], [20]). Applications in the field of biology have ranged from the spread of diseases to models for genomics ([35], [29], [37] [27], [6]). In social sciences spatio-temporal models can be used to predict the behavior and response of individuals and groups ([16]). An abundance of further examples and references for spatio-temporal analyses can be found in [8].

In general, there are two main modeling approaches to spatio-temporal problems: those driven by the underlying physics of the system (deterministic) and those driven by data (stochastic). Examples of physically driven systems are models for producing numerical weather predictions ([33]), describing the dynamics of a boiler ([3]) and modeling mud flow down a slope ([11]). These types of models have their roots in a set of physical laws or local behavior such as the conservation of momentum or mass and energy balances. Examples of data-driven models include modeling space-time scenarios of wind power generation ([32]), mapping of disease rates ([37]) and modeling of sea surface temperature ([17]). These approaches share the emphasis on the aggregate system behavior as opposed to its local dynamics and tackle the modeling task by placing the focus on probability distributions, correlations and inter-dependences.

This article adds to the literature on spatio-temporal modeling by providing a data-driven approach to capture the local dynamics of a larger system. Specifically, we initially propose a model for the local dynamics using coupled stochastic differential equations. This model is, in turn, generalized to govern all local behavior, thus yielding a global model. Furthermore, this global model can be interpreted as a discretization of a more general model in space and time given by a system of partial stochastic differential equations.

1.1. A Motivating Example

The practical application that motivates our methodology is that of forecasting the power generated by a solar power plant. Our claim is that, by understanding and capturing the specific spatio-temporal dynamics of the solar field, we will be able to improve our power predictions. The power output of a solar plant exhibits dynamics on several different time scales: from cycles spanning a year governed by the sun height and climate, to dynamics in the range of weeks and days governed by weather fronts, to hourly dynamics governed by local weather phenomena and different sun height during the day, to the very short horizon of minutes and seconds as a result of the movement of individual clouds. In this example we are motivated by nowcasting (forecasting in the range from seconds to minutes) the power output of a solar plant by tracking the movement of individual clouds across the solar field.

Our data pertain to the Copper Mountain Solar 1 Facility with a rated capacity of 58 MW, and part of the Copper Mountain Solar Facility with at total rated capacity of 150 MW. Our measurements stem from 96 photovoltaic inverters, each with measurements taken every second. We limit the study to the rectangular grid of 5 by 14 inverters shown in red in Figure 1. This specific cutout data is also used in the paper by [18].

Good forecasts for power production are extremely important for secure, reliable, and efficient operation of the electrical grid [25, 23]. Forecasting the power output dynamics on very short horizons has proven essential for an efficient integration of solar power. On partly cloudy days the power output of a solar plant can drop from nominal capacity to between 20-25% of nominal capacity within just a minute as a cloud passes overhead. With a nominal capacity of 150 MW, a drop of 75% results in a power output decrease of 112.5 MW, which may severely challenge grid stability. Very short-term forecasts may help mitigate the detrimental effects of this power drop by either installing storage devices or by providing an early warning for grid operators. Spatio-temporal forecasts of solar irradiance have been an output of physical models such as numerical weather prediction models for decades ([33]). However, besides being deterministic in nature, numerical weather prediction models do not provide enough resolution in space or time to track clouds. Spatio-temporal models for solar power production have just recently received attention: [39] formulate an auto-

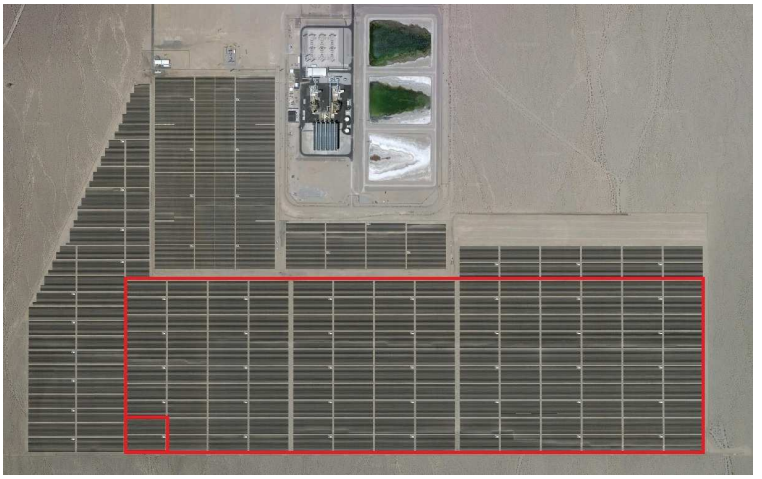

Figure 1: *An aerial photo of the Copper Mountain Solar Facility. The small red rectangle indicates the size of a single inverter, which is 125m by 125m. The large red rectangle indicates the 5 by 14 inverters for which we provide forecasts of the power production.*

regressive time-series model for capturing the correlation in the solar power output across a small area. In [40] space-time kriging and a vector auto-regressive model are employed to describe the spatio-temporal solar power production. [18] outperform a persistence model for short horizons by means of a cloud speed persistence model that propagates the solar power production across the spatial grid in the direction of the cloud speed.

Using stochastic differential equations to model physical systems with a large random component is not new. However, the coupling and modeling methodology presented here is a novel combination and it constitutes a step forward for spatio-temporal modeling in terms of computational efficiency, a low-dimension parameter space, predictions for multiple horizons and characterization of the spatio-temporal interdependence. Furthermore, for the specific case of solar power forecasting, the proposed modeling approach allows for a particularly elegant interpretation of the global system dynamics. The rest of the paper is structured as follows: Section 2 introduces the proposed stochastic differential equation framework. Section 3 describes the approach for using this framework to develop spatio-

temporal models in general and for the specific application to solar power forecasting. In Section 4 we obtain a generalized interpretation of the proposed model. Section 5 details the parameter estimation procedure. We then evaluate the performance of the obtained spatio-temporal model on a real-world example in Section 6. Lastly, Section 7 concludes the paper.

2. Stochastic Differential Equations

Stochastic differential equations (SDEs) are an extension of ordinary differential equations (ODEs) obtained by including one or more stochastic terms. The solution to a SDE is a stochastic process describing the evolution of a random variable over time. SDEs have been used to describe a variety of phenomena governed by a large random component and are especially prominent in finance ([4], [19]) and physics ([34], [1]). We give here a very short introduction to SDEs and refer the interested reader to [24] for a complete and thorough treatment of the subject.

Suppose that we have a continuous time process $U_t \in \mathcal{U} \subset \mathbb{R}^n$. From ordinary differential equations the evolution in time of the state variable, U_t, is defined by the deterministic system equation:

$$\frac{dU_t}{dt} = f(U_t, t), \tag{1}$$

where $t \in \mathbb{R}$ and $f(\cdot) \in \mathbb{R}^n$. For complex systems the dynamics may be too intricate to be captured fully by $f(\cdot)$ or there may be random perturbations of inputs that are not specified by the model. This suggests the introduction of a random component in the state evolution to capture such perturbations or model deficiencies. By introducing a random component in the dynamics of the state process, as carried out in [24], we obtain the following state process:

$$\frac{dU_t}{dt} = f(U_t, t) + g(U_t, t)W_t, \tag{2}$$

where $W_t \in \mathbb{R}^m$ is an m-dimensional standard Wiener process and $g(\cdot) \in \mathbb{R}^{n \times m}$ is a matrix function. Multiply by dt on both sides of (2) to obtain the standard SDE formulation:

$$dU_t = f(U_t, t)dt + g(U_t, t)dW_t. \tag{3}$$

This standard formulation for SDEs is not well defined, as the derivative of W_t, $\frac{dW_t}{dt}$, does not exist. Instead, equation (3) should be interpreted as an informal way of writing the integral equation:

$$U_t = U_0 + \int_0^t f(U_t, t)dt + \int_0^t g(U_t, t)dW_s. \tag{4}$$

In equation (4) the behavior of the stochastic process U_t is expressed as the sum of an initial stochastic variable, a Lebesgue integral and an Itō integral, respectively.

In general, it is only feasible to observe a continuous time process in discrete time. To this end, we observe the process U_t at discrete times through an observation equation. Let $Y_k \in \mathcal{Y} \subset \mathbb{R}^l$ denote the observation at the discrete time t_k. We define the observation equation as:

$$Y_k = h\left(U_{t_k}, t_k, e_k\right), \tag{5}$$

where the introduction of t_k allows for some external input, $e_k \in \mathbb{R}^l$ is the observation error and the function $h(\cdot) \in \mathbb{R}^l$ links the process state to the observation.

The solution to a deterministic ordinary differential equation is a point for each future time t. In the SDE setting the solution is a stochastic process with a probability density for any state and for any future time t. For an Itō process defined as in (4) with drift $f(U_t, t)$ and $g(U_t, t) = \sqrt{2D(U_t, t)}$, the probability density function $p(u, t)$ in state u at time t of the random variable U_t is given as the solution to the partial differential equation known as the Fokker-Planck equation ([4]):

$$\frac{\partial}{\partial t}p(u, t) = -\frac{\partial}{\partial u}\left[f(u, t)p(u, t)\right] + \frac{\partial^2}{\partial u^2}\left[D(u, t)p(u, t)\right]. \tag{6}$$

Thus, given a specific SDE formulation, the predictive density for any future time can be obtained by solving a partial differential equation. While analytic solutions only exist for particularly simple SDE formulations, a host of numerical solutions are available ([14], [31]). While not solving the Fokker-Planck equation directly, another technique, the Monte Carlo approach, solves this problem implicitly by approximating the predictive density through simulation ([28]).

3. A Spatio-Temporal Model by Coupled SDEs

Consider a stochastic process in space x and time t, and denote this process by $U(x,t)$. Suppose that there are $I \times J$ locations at $x_{i,j}$ where we want to model the process. First, let $U(x,t)$ at location $x_{i,j}$ be denoted by $U(x_{i,j},t) = U_{i,j,t}$. Now suppose that we want to use a stochastic differential equation to represent the dynamics of each $U_{i,j,t}$. This gives us a model of the following form:

$$dU_{i,j,t} = f(\mathbf{U}_t, t)\,dt + g(\mathbf{U}_t, t)dW_{i,j,t} \quad (7)$$

$$Y_{l,k} = h(\mathbf{U}_{t_k}, t_k) + \epsilon_{l,k}, \quad (8)$$

where we let $\mathbf{U}_t$ be the vector containing all $U_{i,j,t}$ for a specific t.

Next, we enforce that two locational processes have a direct interaction only if they are adjacent to each other. This stems from a physical interpretation of the system, whereby we allow only locations that are in direct contact to interact with each other. To this end, define $\mathbf{U}_{i,j,t}$ as the set of $U_{i,j,t}$'s that are in the nearest neighborhood of $U_{i,j,t}$, thus $\mathbf{U}_{i,j,t} = \{U_{i,j,t}, U_{i-1,j,t}, U_{i+1,j,t}, U_{i,j-1,t}, U_{i,j+1,t}\}$. Furthermore, assume that for each $U_{i,j,t}$ we allow $g(\cdot)$ to depend only on $U_{i,j,t}$ and t. This also follows from a physical interpretation of the system dynamics, whereby random perturbations at location $x_{i,j}$ can only affect adjacent locations by first affecting $U_{i,j,t}$. This leads to the following model formulation:

$$dU_{i,j,t} = f(\mathbf{U}_{i,j,t}, t)\,dt + g(U_{i,j,t}, t)dW_{i,j,t} \quad (9)$$

$$Y_{l,k} = h(\mathbf{U}_{t_k}, t_k) + \epsilon_{l,k}. \quad (10)$$

Since the model is formulated in continuous time, by appealing to the physical nature of the system, one should expect there to be no interaction between locations that are not adjacent to each other. The model formulation is illustrated in Figure 2. Here we have depicted the different locations by dots and interactions with lines connecting two dots. Specifically we have highlighted in red all interactions that concern location $U_{i,j,t}$. This specific formulation is defined for 2-dimensions but can be easily generalized to higher dimensions.

3.1. Application to Solar Power Forecasting

Based on the model formulation outlined in this section, we propose a model for predicting the power output of a photovoltaic solar power facility. The model exploits the power

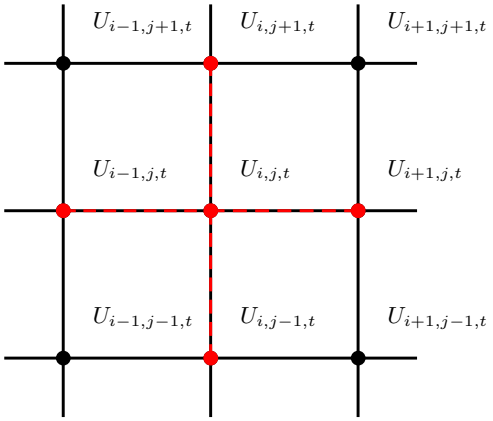

Figure 2: *A stencil showing the relationship between $U(x,t)$ at different locations, with black lines indicating a direct relationship. The red dashed lines show the interactions for $U_{i,j,t}$*

output of the up-wind solar inverters to predict the future power output of the down-wind solar inverters.

The data available pertains to the Sempra US Gas & Power Copper Mountain Solar Facility outside Boulder City, Nevada, USA. The power output of the solar power inverters is normalized both with respect to the sun height and with respect to the solar panel tilt and we henceforth refer to the normalized power as power. We let the *change* in power output of inverter $[i,j]$ at location $x_{i,j}$ at time t be modeled by the stochastic variable $U_{i,j,t}$. We order the inverters such that inverter $U_{i+1,j,t}$ is the one directly to the east of inverter $U_{i,j,t}$. Also, we name the inverters such that $U_{i,j+1,t}$ is the inverter directly north of $U_{i,j,t}$ (see Figure 2).

We have cloud speed measurements at our disposal, provided by the approach given in [5]. The cloud speed is denoted by v_t. For modeling purposes, the cloud speed is decomposed into its four directional components, namely, North, East, South, and West, denoted by n_t, e_t, s_t, and w_t, respectively. We employ four directional components instead of two, as we impose the condition that the directional component must be non-negative. Thus, a wind from South-East has positive South and East components, but zero North

and West components. This way we end up with the model:

$$\begin{aligned} dU_{i,j,t} &= \theta|v_t|\Big(\mathrm{n}_t(U_{i,j+1,t}-U_{i,j,t})\mathbb{1}_{\{j+1\leq J\}}+\mathrm{e}_t(U_{i+1,j,t}-U_{i,j,t})\mathbb{1}_{\{i+1\leq I\}} \\ &\quad +\mathrm{s}_t(U_{i,j-1,t}-U_{i,j,t})\mathbb{1}_{\{j-1\geq 1\}}+\mathrm{w}_t(U_{i-1,j,t}-U_{i,j,t})\mathbb{1}_{\{i-1\geq 1\}} \\ &\quad -\mu U_{i,j,t}\left(\mathrm{n}_t\mathbb{1}_{\{j=J\}}+\mathrm{s}_t\mathbb{1}_{\{j=1\}}+\mathrm{e}_t\mathbb{1}_{\{i=I\}}+\mathrm{w}_t\mathbb{1}_{\{i=1\}}\right)\Big)dt \\ &\quad +\sigma dW_{i,j,t} \end{aligned} \tag{11}$$

$$dQ_{i,j,t} = U_{i,j,t}dt \tag{12}$$

$$Y_{i,j,k} = Q_{i,j,t_k}+\epsilon_{i,j,k}. \tag{13}$$

Here $Y_{i,j,k}$ is the observed power produced from location $x_{i,j}$ at time t_k. $Q_{i,j,t}$ can thus be interpreted as the actual produced power at this location. This implies that $U_{i,j,t}$ is the change in power production at location $x_{i,j}$ at time t. Notice that the spatial dynamics are modeled by equation (12) and that equations (12)–(13) correspond to the observation equation (10). Thus, to express the SDE model (12)–(13) in the form of (7)–(8), it suffices to write the $h(\cdot)$ function as $h_{i,j}(t_k)=\int_{t_{k-1}}^{t_k}U_{i,j,s}ds+\epsilon_{i,j,k}$. Further we have that $\epsilon_{i,j,k}\sim\mathcal{N}(0,\sigma_\epsilon^2)$. The parameters in the model are thus θ,μ,σ and σ_ϵ, where $\theta|v_t|$ governs the speed at which the value in adjacent cells tend towards each other. Parameter μ governs how rapidly the change in power output of inverter $\{i,j\}$, $U_{i,j,t}$, tends to zero, if $U_{i,j,t}$ is an upwind cell. σ is the system noise and σ_ϵ characterizes the observation noise. Symbol $\mathbb{1}_{\{\cdot\}}$ represents an indicator or heavyside function, that is equal to 1 if the stated condition is met, and 0 otherwise. The indicator functions are used to handle the boundaries of the solar field, such that the model only relates locations that are actually present in the model. This also applies to the dampening term, where we dampen cells on the leading edge towards the wind. Further, note that, in this particular case, $Y_{l,k}=Y_{i,j,k}$, where we let l go through all the feasible combinations of $\{i,j\}$ to conform with the notation in equation (8) (that is, model (12)–(13) assumes that we have power measurements for all locations or inverters).

4. Continuous Space Interpretation

Given the model formulation (12)–(13) one might ask what would happen if the grid size approached zero. Notice that, since we have a fixed distance Δx between all adjacent

grid points, we can normalize the model parameters by doing $\theta = \tilde{\theta}/\Delta x$, $\sigma = \tilde{\sigma}\Delta x$ and $\mu = \tilde{\mu}\Delta x$, with $\Delta x = c$, where c is some constant. Consequently, the SDE model (12)–(13) can be recast as:

$$\begin{aligned} dU_{i,j,t} &= \tilde{\theta}|v_t| \left(\mathrm{n}_t \left(\frac{U_{i,j+1,t} - U_{i,j,t}}{\Delta x} \right) \mathbb{1}_{\{j+1 \leq J\}} + \mathrm{e}_t \left(\frac{U_{i+1,j,t} - U_{i,j,t}}{\Delta x} \right) \mathbb{1}_{\{i+1 \leq I\}} \right. \\ &\quad + \mathrm{s}_t \left(\frac{U_{i,j-1,t} - U_{i,j,t}}{\Delta x} \right) \mathbb{1}_{\{j-1 \geq 1\}} + \mathrm{w}_t \left(\frac{U_{i-1,j,t} - U_{i,j,t}}{\Delta x} \right) \mathbb{1}_{\{i-1 \geq 1\}} \qquad (14) \\ &\quad \left. - \tilde{\mu} U_{i,j,t} \left(\mathrm{n}_t \mathbb{1}_{\{j=J\}} + \mathrm{s}_t \mathbb{1}_{\{j=1\}} + \mathrm{e}_t \mathbb{1}_{\{i=I\}} + \mathrm{w}_t \mathbb{1}_{\{i=1\}} \right) \right) dt \qquad (15) \\ &\quad + \tilde{\sigma} \Delta x dW_{i,j,t} \\ dQ_{i,j,t} &= U_{i,j,t} dt \qquad (16) \\ Y_{i,j,k} &= Q_{i,j,t_k} + \epsilon_{i,j,k}, \qquad (17) \end{aligned}$$

Now notice that for the easterly direction

$$\lim_{\Delta x \to 0} \frac{U_{i+1,j,t} - U_{i,j,t}}{\Delta x} = \left(\frac{\partial U(x,t)}{\partial x_1} \right)_{x_{i,j}}. \qquad (18)$$

Similarly, we can compute this quantity for the other directions.

Next, consider the integral:

$$\int_t^{t+\Delta t} \Delta x dW_{i,j,t} \sim \mathcal{N}(0, \Delta x^2 \Delta t) \qquad \forall \quad \Delta x, \Delta t \geq 0. \qquad (19)$$

This is exactly equal to the definition of a Brownian motion in 2D space and time.

Hence it becomes evident that, when we look at the set of coupled stochastic differential equations given by (12) away from the boundaries, this set can be interpreted as a finite difference discretization of the following stochastic partial differential equation (SPDE):

$$dU(x,t) = \bar{v}\theta \nabla U(x,t) dt + \sigma dW(x,t), \qquad (20)$$

where $W(x,t)$ is a Brownian motion in space and time, $\bar{v}$ is the cloud speed vector, and ∇ is the partial derivative operator. Other SPDE models have originated following an analogous micro-to-macro approach ([2], [10]).

Some intuition about the SPDE (20) can be gained by looking at its deterministic part, namely:

$$dU(x,t) = \bar{v}\theta \nabla U(x,t) dt, \qquad (21)$$

which is a uni-direction wave equation, where the direction is determined by the cloud speed vector.

5. Fitting Procedure

The estimation of parameters is carried out using the statistical software R ([26]) and in particular the package CTSM-R (Continuous Time Stochastic Modeling for R) ([15]). The method is based on the Kalman filter for obtaining the likelihood. We provide a brief overview of the approach implemented and refer the reader to [15] for more details on the approach.

The discretized SPDE (14)-(17) is linear in the states $U_{i,j}$ and thus can be formulated as a linear SDE with linear observations

$$dU_{i,j,t} = \mathbf{A}(\theta,t)\mathbf{U}_t dt + \Sigma(\theta) dW_{i,j,t} \tag{22}$$

$$Y_k = \mathbf{C}(\theta)\mathbf{U}_{t_k} + \epsilon_k, \tag{23}$$

where $\mathbf{A}(\theta,t)$ is a time varying transition matrix, $\Sigma(\theta)$ is the diffusion matrix, $\mathbf{C}(\theta)$ determines how the states are observed and $e_k \sim \mathcal{N}(0,\sigma^2)$. The aim is to estimate the parameter vector θ in the model defined by the linear equations (22) - (23). The likelihood depends only on the one-step ahead prediction probability densities of the observations. A linear model driven by Gaussian diffusion results in a Gaussian process which is fully described by the mean and variance of the observations. We define them as

$$\widehat{Y}_{k|k-1} = \mathbb{E}\left[Y_k|\mathcal{Y}_{k-1},\theta\right] \tag{24}$$

$$R_{k|k-1} = \mathbb{V}\left[Y_k|\mathcal{Y}_{k-1},\theta\right], \tag{25}$$

where $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ denote the expectation and variance, respectively, and $\mathcal{Y}_{k-1} = \{Y_0,\ldots,Y_{k-1}\}$. We can now define the innovation error as the difference between the observed and expected outcome:

$$\epsilon_k = Y_k - \widehat{Y}_{k|k-1}, \tag{26}$$

which will be used to compute the likelihood. We require in equations (7) - (7) that $g(U_t,t) = g(t)$ and $h(U_{t_k},t_k,e_k) = h(U_{t_k},t_k) + e_k$, where $e_k \sim \mathcal{N}(0,\sigma^2)$. However these

requirements can be alleviated to a large extent through transformations of the state equations ([12], [21]) or transformations of the observations ([7]). For a system satisfying these conditions, the likelihood is given by

$$L\left(\theta;\mathcal{Y}_N\right)=\left(\prod_{k=1}^{N}\frac{\exp\left(-\frac{1}{2}\epsilon_k^\top R_{k|k-1}^{-1}\epsilon_k\right)}{\sqrt{\det\left(R_{k|k-1}\right)}\left(\sqrt{2\pi}\right)^l}\right)p(Y_0|\theta). \tag{27}$$

Here l is the dimension of the sample space, thus the dimension of Y_k, N is the number of observations, $(\cdot)^\top$ denotes the vector transpose and $p(Y_0|\theta)$ is the likelihood of seeing observation Y_0.

We are tracking changes in the observed power output, but by far the majority of the data is without cloud activity. To reduce the computational load, we extract $M = 12$ segments, each of 3 hours each with power measurements every 5 seconds to reduce the burden of the estimation process. The M sets are from separate days spread out such that we have a sample day from each month of the year. Thus, the data sets can be assumed independent. The likelihood of M independent sets of observations is

$$L\left(\theta;\mathbf{Y}\right)=\prod_{i=1}^{M}\left(\prod_{k=1}^{N_i}\frac{\exp\left(-\frac{1}{2}{\epsilon_k^i}^\top {R_{k|k-1}^i}^{-1}\epsilon_k^i\right)}{\sqrt{\det\left(R_{k|k-1}^i\right)}\left(\sqrt{2\pi}\right)^l}\right)p(Y_0^i|\theta), \tag{28}$$

where $\mathbf{Y}=\left[\mathcal{Y}_{N_1}^1,\mathcal{Y}_{N_2}^2,\ldots,\mathcal{Y}_{N_M}^M\right]$ is the combined set of observations and N_i is the number of observations in each data set. We consider the logarithm of the likelihood function conditional on $\mathbf{Y}_0=\left[Y_0^1,Y_0^2,\ldots,Y_0^M\right]$, both for computational considerations and in order to deal with the fact that there are no observations prior to Y_0. This results in:

$$\begin{aligned}\log\left(L\left(\theta;\mathbf{Y}_N|\,Y_0\right)\right) &= -\frac{1}{2}\sum_{i=1}^{M}\sum_{k=1}^{N_i}\left(\log(\det(R_{k|k-1}^i))+{\epsilon_k^i}^\top {R_{k|k-1}^i}^{-1}\epsilon_k^i\right)\\ &\quad -\log(2\pi)\frac{l\sum_{i=1}^{M}N_i}{2}.\end{aligned} \tag{29}$$

The parameter vector θ enters the log-likelihood function (29) through ϵ_k^i and $R_{k|k-1}^i$. An estimate of the parameters in the model can now be obtained by maximizing (29), i.e.,

$$\hat{\theta}=\arg\max_{\theta\in\Theta}\left(\log\left(L\left(\theta;\mathbf{Y}_N|\,\mathbf{Y}_0\right)\right)\right), \tag{30}$$

where Θ is the feasible parameter space. A thorough introduction to parameter estimation and filtering is found in [13]. We note that the likelihood function is optimized for the one-step-ahead residuals. To estimate SDE models using a multi-horizon approach, we refer the interested reader to [22].

5.1. Speeding up the estimation process

We use a large amount of data during the estimation procedure which naturally slows down the computation. Since the likelihood of the 12 independent sets of observations is simply the product of the likelihood of each of the sets, we can evaluate these likelihoods in parallel. Furthermore, the optimization problem (30) is solved using a quasi Newton algorithm, where the gradient of the log likelihood (29) is determined by a finite difference scheme. This means evaluating (29) at several independent points in the parameter space Θ. Thus, the computation of the gradient can also be parallelized.

A server with 2x12 cores AMD Opteron 6168 CPUs was used for estimation and prediction. We use OpenMP and nested parallelism to maximize the use of the server. The log likelihood (29) is always computed in parallel using 12 threads. When computing the gradient, two evaluations of the log likelihood were allowed simultaneously, and in so doing all the available 24 cores were used. We achieved a total speedup of 15.6x.

6. Nowcasting at Copper Mountain First Solar

In this section we look at the specific problem of nowcasting the power output of the Copper Mountain Solar facility. We fit the model on a training data set consisting of data from 12 days, one from each month of the year, where we select 3 hours around noon. We select only days where there are actually observed clouds as we propose to model the cloud dynamics. Similarly, we select a test data-set not overlapping with the training data set. We first run the estimation procedure described in Section 5 on the model described in Section 3.1. As this is computationally intensive, we estimate the parameters on a cutout of 5 by 7 inverters to allow for a timely estimation procedure. The parameters found here are then used to define the full model spanning 5 by 14 inverters. The parameter estimates obtained are shown in Table 1.

$\hat{\theta}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\sigma}_\epsilon$
0.0631	0.703	0.00865	10^{-10}

Table 1: *Parameter estimates for the proposed model*

Figure 3 displays the actual power output of the inverters along with the predicted power for different horizons. It can be observed that we manage to track cloud movement through the system as regions with lower power output. However, for longer horizons the predicted power may vary to a large degree from the observed power. This can be explained by several facts: first, that the clouds that caused the drop in power were not observed by the upwind inverters at the time when the forecast was issued. Second, there might be some smaller errors in the estimation of the cloud speed vector. These errors compound to produce larger errors for larger lead times. Third, looking at the observations in the left panels, it seems that the actual structure of the cloud actually changes over time. This may, in part, be due to the spatial resolution of the observations and, in part, due to real changes of the cloud structure.

In Figure 4 the normalized total power output of the solar field is shown together with predictions issued for different lead times. We see that for 5- and 20-second horizons the model seems to be successful at predicting the output power. For longer horizons we can observe a "lagging" behavior. This is caused by the fact that the clouds causing the power drop (or increase) had not yet begun to enter into the system (leave the system) and therefore, to be detected. As a result, their future effects are not anticipated by the forecasts. Furthermore, the predictions become less smooth as we predict for longer lead times. This is caused by the predictions for total power being based on fewer actual observations, since the influence of many inverters is propagated out of the system. There is also an error propagation, where small errors accumulate over time.

Second, we compare the model proposed in this paper with state-of-the-art models for spatio-temporal solar power forecasting. We compare the different models on several horizons to better understand the specific characteristics of each . The benchmarks include a cloud speed persistence model (as defined in [18]), which propagates the power production along the cloud speed vector. Another benchmark is the ramp speed persistence model (also

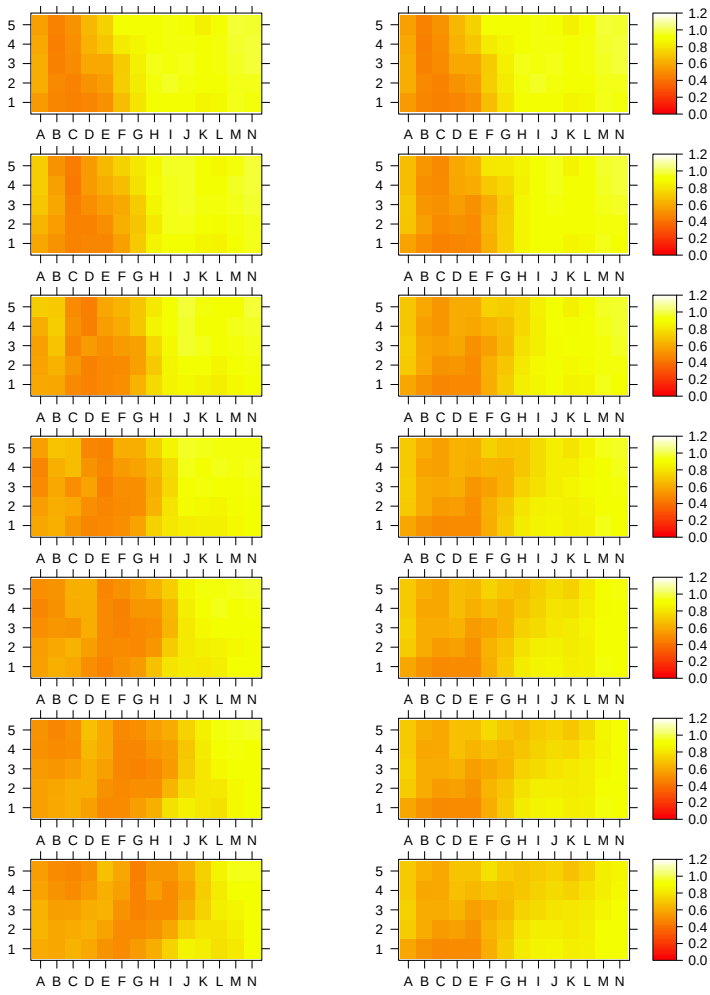

Figure 3: *The observed power generation (left) and the predicted power generation (right) from 0 to 60 seconds in 10 second increments. The x-axis corresponds to the East-West axis and the y-axis corresponds to the North-South axis.*

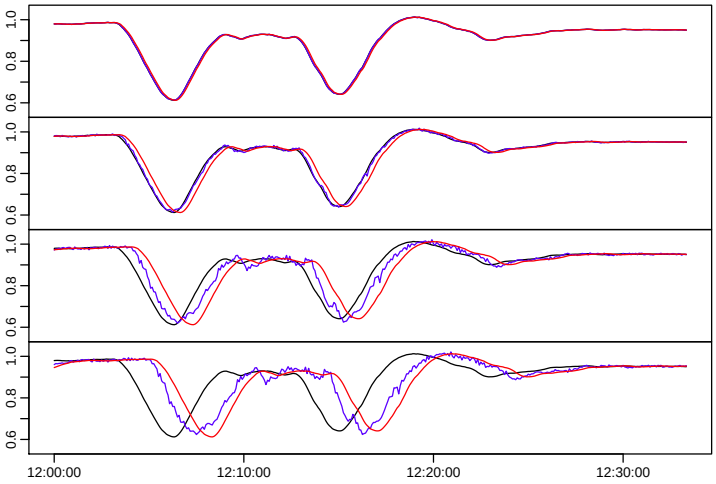

Figure 4: *The observed normalized total power in black, along with the (from top to bottom) 5, 20, 60 and 120 sec ahead forecasts from the model in blue and the persistence forecast in red. The x-axis is time in hours, minutes and seconds, the y-axis is normalized power.*

in [18]), where the change in power is assumed to stay constant for the near-term future. A third benchmark is an auto-regressive model defined as:

$$Y_k = \psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \epsilon_k, \qquad \text{where } \epsilon_k \sim \mathcal{N}(0, \sigma^2), \tag{31}$$

ψ_i are the auto-regressive parameters of the model and p defines the number of lags included.

In Table 2 the proposed model is compared with the benchmarks in terms of skill scores against the persistence benchmark for the root mean squared error (RMSE) and the mean absolute error (MAE) of total power production. The persistence is the lagged value of the observations, here the lag is given as the forecast horizon. This skill score is computed as:

$$\text{SS} = 1 - \frac{S_{\text{forecast}}}{S_{\text{ref}}}, \tag{32}$$

where SS is the skill score for the forecast score, S_{forecast}, against the reference score, S_{ref} , obtained from the reference model (In this application the reference model is the persistence).

For benchmarks that produce predictive densities, we compare in terms of continuous ranked probability score (CRPS) (as defined in [9]) to evaluate the probabilistic properties of the predictions.

Score	Cloud Speed Persistence	Ramp Speed Persistence	Auto-Regressive	Model (12)-(13)
$RMSE_5$	0.334	0.612	0.464	**0.636**
$RMSE_{20}$	0.289	0.284	0.319	**0.523**
$RMSE_{60}$	0.168	-0.203	0.113	**0.254**
$RMSE_{120}$	0.062	-0.434	0.039	**0.097**
MAE_5	0.258	0.597	0.431	**0.612**
MAE_{20}	0.213	0.301	0.280	**0.497**
MAE_{60}	0.136	-0.145	0.045	**0.246**
MAE_{120}	0.048	-0.396	-0.064	**0.096**
$CRPS_5$	–	–	0.00262	**0.00131**
$CRPS_{20}$	–	–	0.00982	**0.00666**
$CRPS_{60}$	–	–	0.02886	**0.02455**
$CRPS_{120}$	–	–	0.04883	**0.04675**

Table 2: *The MAE skill score, RMSE skill score and CRPS for benchmarks as well as for the proposed model for horizons of 5, 20, 60 and 120 seconds.*

In Table 2 we see that the coupled SDE model (12)–(13) outperforms all benchmarks on all horizons in terms of all the proposed scores.

The scores in Table 2 are computed on the basis of total output power. However, as the proposed model also captures the dynamics of each individual inverter, we might well evaluate the predictive performance of the individual inverters. This is done with respect to the RMSE skill score for 20 seconds ahead in Figure 5. As this is a skill-score, higher score values are better. As seen in Figure 5 the inverters that perform the poorest are located on the southern and western limits of the solar plant. Investigating this phenomenon we find that prevailing winds are south-westerly. We would expect the up-wind inverters to perform worse compared to down wind inverters due to the influx of clouds. Thus the

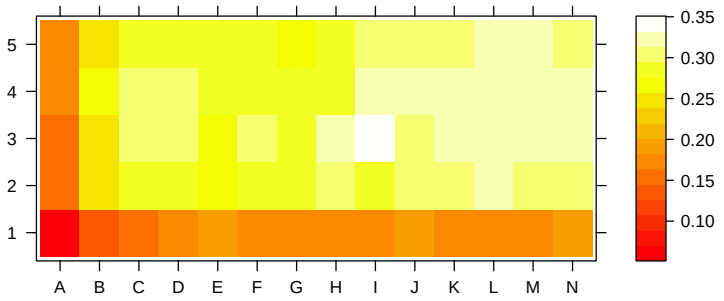

Figure 5: *The RMSE skill score for 20 seconds ahead forecasts computed for individual inverters.*

findings from our model are in accordance with what would be expected.

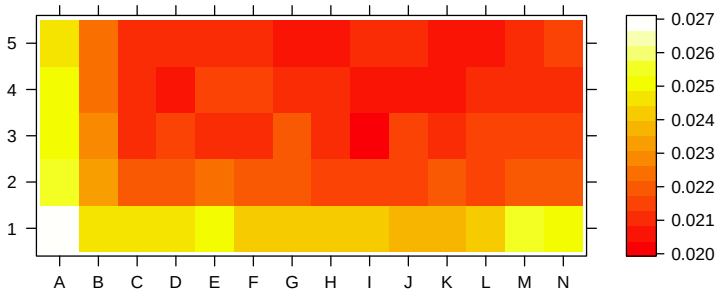

Figure 6: *The CRPS for 20 seconds ahead forecasts computed for individual inverters. The x-axis corresponds to the East-West axis and the y-axis corresponds to the North-South axis.*

Figure 6 is analogous to Figure 5 but in terms of the CRPS. As opposed to Figure 5, lower CRPS values are better. Again we see the better performance in the interior of the solar power plant.

The proposed spatio-temporal model outperforms persistence as well as all the proposed state-of-the-art benchmarks. A spatial understanding of the dynamics not only allows for spatio-temporal predictions, but also allows us to better predict the aggregate power production. Furthermore, we note that the performance of the benchmarks that are used

here is similar to the performance of the benchmarks found in [18]. It is crucial to stress the importance of the inputs into the model being the cloud speed vector obtained through the approach in [5]. This cloud speed vector has crucial importance for the correct propagation of the irradiance field. As mentioned in [5], estimating the correct cloud speed vector is not a simple task. This is, in part, due to the granularity of the spatial observations and, in part, due to the deformation of the clouds. Even small errors in the cloud speed vector lead to serious errors when the forecast horizon is increased. In [5] it is also clearly stated that there is a large degree of uncertainty related to this clouds speed vector.

7. Concluding Remarks

Spatio-temporal problems arise in many fields ranging from physics to ecology. The processes that drive these systems can be quite complicated. However, there are often scientific theories that explain the local behavior of the system, e.g., mass and energy balances or the migration of animals. Although, generally, traditional statistical methods are not well-suited to model such processes, the modeling framework that we propose in this paper, based on coupled stochastic differential equations, proves to perform satisfactorily. Coupled stochastic differential equations have the capacity to capture dynamics where the structure of the governing stochastic partial differential equation is not well known and they can be used to identify a possible candidate, as we showed here. Furthermore, this approach bridges the gap between spatio-temporal models that are driven by physics and those which are driven by data. A key feature is that the model framework reduces the parameter dimension of the spatio-temporal problem and thereby facilitates parameter estimation and efficient computation. In this paper the methodology for building spatio-temporal models is applied to forecasting solar power generation at a solar power facility.

The spatio-temporal forecast model proposed in this paper outperforms state-of-the-art benchmarks on all horizons while also being able to provide scenarios, covariances and predictive densities. Understanding the spatial dynamics not only allows for spatio-temporal predictions but also allows us to produce higher quality predictions for aggregate power production. The model generates predictions swiftly and as such, could be run online. Thus, we produce a methodology for predicting large ramp events 30-60 seconds

ahead in time (depending on cloud speed and direction) providing forecast users with an early warning to scramble alternative power generation.

The application to solar power forecasting that we introduce in this paper assumes the same local dynamics for every location. This is not, however, a requirement. Distinct inputs could be present for each specific location, distinguishing the dynamics at different grid points. Also, if this framework were to be applied to forecast power generation from distributed solar power in large urban areas, the grid would change and there may be distinct features in space to consider. A clear conclusion from the results shown in Figure 6 is that the model could be extended to have irradiance sensors away from the solar power facility in order to increase forecast performance and to extend the forecast horizon.

The spatio-temporal model considered in this paper is of a particularly simple structure, with regular grid spacing in all spatial dimensions. Nonetheless, this is not a prerequisite for applying a similar model to a non-regular grid and as such, more research efforts can be dedicated at adapting this approach to irregular grids. Future work could also include using the sparse structure of the coupled stochastic differential equations to further reduce computational time. This sparsity is caused by the very nature of the approach, where only locations that are adjacent to each other interact.

Acknowledgments

DSF (Det Strategiske Forskningsråd) is to be acknowledged for partly funding the work of Emil B. Iversen, Juan M. Morales and Henrik Madsen through the Ensymora project (no. 10-093904/DSF). Furthermore, Juan M. Morales and Henrik Madsen are partly funded by the Research Centre CITIES (no. 1035-00027B), which is also supported by DSF (Det Strategiske Forskningsråd). The work was completed as part of a research stay at the University of California, San Diego, made by Emil B. Iversen.

References

[1] Adomian, G. (1988). *Nonlinear Stochastic Systems Theory and Application to Physics*, volume 46. Springer.

[2] Allen, E. J. (2008). Derivation of stochastic partial differential equations. *Stochastic Analysis and Applications*, 26(2):357–378.

[3] Åström, K. J. and Bell, R. D. (2000). Drum-boiler dynamics. *Automatica*, 36(3):363–378.

[4] Björk, T. (2009). *Arbitrage Theory in Continuous Time*. Oxford Finance Series. OUP Oxford.

[5] Bosch, J. and Kleissl, J. (2013). Cloud motion vectors from a network of ground sensors in a solar power plant. *Solar Energy*, 95:13–20.

[6] Bowman, F. D. (2007). Spatiotemporal models for region of interest analyses of functional neuroimaging data. *Journal of the American Statistical Association*, 102(478):442–453.

[7] Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252.

[8] Cressie, N. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons.

[9] Gneiting, T. and Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

[10] Hairer, M. (2009). An introduction to stochastic PDEs. *arXiv preprint arXiv:0907.4178*.

[11] Huang, X. and Garcia, M. H. (1998). A Herschel–Bulkley model for mud flow down a slope. *Journal of fluid mechanics*, 374:305–333.

[12] Iversen, E. B., Morales, J. M., Møller, J. K., and Madsen, H. (2014). Probabilistic Forecasts of Solar Irradiance by Stochastic Differential Equations. *Environmetrics*, 25(3):152–164.

[13] Jazwinski, A. H. (2007). *Stochastic Processes and Filtering Theory*. Courier Dover Publications.

[14] Johnson, C. (2012). *Numerical solution of partial differential equations by the finite element method*. Courier Dover Publications.

[15] Juhl, R., Kristensen, N. R., Bacher, P., Kloppenborg, J., and Madsen, H. (2013). CTSM-R user guide. *Technical University of Denmark. Available at* ***http://ctsm.info/pdfs/userguide.pdf***.

[16] Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al. (2009). Life in the network: The coming age of computational social science. *Science*, 323(5915):721.

[17] Lemos, R. T. and Sansó, B. (2009). A spatio-temporal model for mean, anomaly, and trend fields of North Atlantic sea surface temperature. *Journal of the American Statistical Association*, 104(485):5–18.

[18] Lipperheide, M., Bosch, J., and Kleissl, J. (2015). Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. *Solar Energy*, 112:232–238.

[19] Mikosch, T. (1998). *Elementary Stochastic Calculus: With Finance in View*, volume 6. World Scientific.

[20] Möller, A., Lenkoski, A., and Thorarinsdottir, T. L. (2013). Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society*, 139(673):982–991.

[21] Møller, J. and Madsen, H. (2010). From state dependent diffusion to constant diffusion in stochastic differential equations by the Lamperti transform. Technical report, Technical University of Denmark.

[22] Møller, J., Pinson, P., and Madsen, H. (2013). Probabilistic forecasts of wind power generation by stochastic differential equation models. Technical report, Technical University of Denmark.

[23] Morales, J. M., Conejo, A. J., Madsen, H., Pinson, P., and Zugno, M. (2014). *Integrating Renewables in Electricity Markets – Operational Problems*, volume 205 of *International Series in Operations Research & Management Science*. Springer, New York.

[24] Øksendal, B. (2010). *Stochastic Differential Equations: An Introduction with Applications*. Universitext (1979). Springer.

[25] Pinson, P. et al. (2013). Wind energy: Forecasting challenges for its operational management. *Statistical Science*, 28(4):564–585.

[26] R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[27] Richardson, S., Bottolo, L., and Rosenthal, J. S. (2010). Bayesian models for sparse regression analysis of high dimensional data. *Bayesian Statistics*, 9:539–569.

[28] Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*, volume 319. Citeseer.

[29] Robertson, C., Nelson, T. A., MacNab, Y. C., and Lawson, A. B. (2010). Review of methods for space–time disease surveillance. *Spatial and Spatio-temporal Epidemiology*, 1(2):105–116.

[30] Sigrist, F., Künsch, H. R., and Stahel, W. A. (2014). Stochastic partial differential equation based modelling of large space–time data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

[31] Smith, G. D. (1965). Numerical solution of partial differential equations.

[32] Tastu, J., Pinson, P., and Madsen, H. (2013). Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix. Technical report, Technical University of Denmark.

[33] Uppala, S. M., Kållberg, P., Simmons, A., Andrae, U., Bechtold, V., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., et al. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612):2961–3012.

[34] Van Kampen, N. G. (1992). *Stochastic Processes in Physics and Chemistry*, volume 1. Access Online via Elsevier.

[35] Van Maanen, A. and Xu, X.-M. (2003). Modelling plant disease epidemics. *European Journal of Plant Pathology*, 109(7):669–682.

[36] Versini, P.-A. (2012). Use of radar rainfall estimates and forecasts to prevent flash flood in real time by using a road inundation warning system. *Journal of Hydrology*, 416:157–170.

[37] Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical association*, 92(438):607–617.

[38] Xu, K., Wikle, C. K., and Fox, N. I. (2005). A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. *Journal of the American Statistical Association*, 100(472):1133–1144.

[39] Yang, C., Thatte, A., and Xie, L. (2015). Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation. *Sustainable Energy, IEEE Transactions on*, 6(1):104–112.

[40] Yang, D., Dong, Z., Reindl, T., Jirutitijaroen, P., and Walsh, W. M. (2014). Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage. *Solar Energy*, 103:550–562.

Paper G

Strategies for Charging Electric Vehicles in the Electricity Market

---

Submitted to *International Journal of Sustainable Energy Planning and Management, 2015.*

Strategies for Charging Electric Vehicles in the Electricity Market

Nina Juul[1a], Giovanni Pantuso[a], Jan Emil Banning Iversen[b], Trine Krogh Boomsma[c]
[a] *DTU Management Engineering, Technical University of Denmark, P.O. Box 49, 4000 Roskilde, Denmark*
[b] *DTU Compute, Technical University of Denmark, Matematiktorvet, 2800 Kgs. Lyngby, Denmark*
[c] *Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen East, Denmark*

Abstract

This paper analyses different charging strategies for a fleet of electric vehicles. Along with increasing the realism of the strategies, the opportunity for acting on the regulating market is also included. Particularly, strategies are chosen from uncontrolled charging through deterministic optimization, to modelling the charging and bidding problem with stochastic programming. We show that all vehicle owners will benefit from acting more intelligently on the energy market. Furthermore, the high value of the stochastic solution shows that, in case the regulating price differs from the expected, the solution to the deterministic problem becomes infeasible.

Abbreviations: EV: Electric Vehicle
Keywords: Electric vehicles, Regulating market, Stochastic programming, Bidding

1 Introduction

Increased focus on electrified transportation has an influence on power systems. Charging and discharging could help power plants to produce in a more steady pace, even though an increased amount of fluctuating renewables are influencing the system. Fortunately, optimising the charging from the power system point of view often corresponds to optimising from the vehicle owners point of view - charging when prices are low (high amount of free power producing capacity) and discharging or stop charging when prices are high (and free power producing capacity is low).

From the vehicle owner's point of view, optimising the charging might include participating in the regulating market, hence, bidding capacities for up- and down regulation. This requires enough battery capacity left for either up- or down regulation, and, thus, has an influence on the planned charging at spot price. However, when planning the charging of the vehicle, the regulating prices are unknown and stochastic.

Optimal bidding into the electricity markets has been focused on in many articles. Within the field of mathematical programming, [4] focuses on optimal sequential bidding in both the Day Ahead market and regulating market, considering uncertainty in prices on regulating market. Other

[1]Corresponding author - e-mail: njua@dtu.dk

examples of bidding models are [7, 15]. These models are all considering price-taking electricity producers and include details such as start-up costs, ramping restrictions, and storage balances.

The charging of electric vehicles (EVs) has been a focus area for a considerable number of articles. This diverges from the challenges and benefits in the entire energy system [11, 16] to optimal charging when driving patterns are stochastic, e.g. [10]. in [14], a deterministic model has been developed, showing incentives for flexible charging. They have clustered the vehicles depending on driving patterns, and show how optimal charging primarily fills the valleys of electricity demand.

A few of the articles have focused on the electric vehicles bidding in the power market. In [2], a deterministic model with hourly time steps has been used to optimise bidding on both the Day Ahead market and for secondary reserves. [6] provides a dynamic approach to the bidding problem, focusing on regulating reserves only. Here, the bidding is split into two time periods; day 8-20 and night 20-8, and each bid counts for an entire period. Furthermore, they argue that large vehicle pools compensate for the stochastic variation.

Stochastic programming has been used by [13, 1, 17, 8]. [13] maximizes the revenue to the aggregator in a two stage model, where bids are places in the first stage and realised in the second. No discharging is allowed and the Day Ahead market is not included. Another two stage model is developed in [1], where they mitigate risk by coordinating bids on Day Ahead market between wind power, thermal power, and electric vehicles. Thereby, they try to minimise the trading risks from market and wind uncertainties. [17] also focus on two-stage problems, where they focus on bidding in both markets simultaneously and bidding in the Day Ahead market only followed by participation on the regulating market. For the latter (regulating market only), they use rolling planning, hence, a series of two stage problems, to optimise for each hour of the day. The principle of rolling planning has been extended in [8] by using a multi-stage model, maximising probability that the regulation bid is accepted. The focus is on plug-in hybrid vehicles, resulting in the vehicles being able to drive even though, the charging does not meet the target.

The scope of this article is to investigate the value of more sophisticated modelling when optimising the charge of electric vehicles. This is done by building a two-stage stochastic model to optimise the charging of a number of electric vehicles with different driving patterns - acting on the regulating market when possible and beneficial. These results are compared to those of a deterministic model and the results of other heuristic charging strategies. Hereby, we are looking into the value of more sophisticated and, thus, time consuming models to be used as decision support tools in respect to charging. Furthermore, we allow for both up and down regulation in terms of either charging when not planned (down) and stopping a planned charging (up). Hence, we do not allow for discharging of the vehicles.

The article is structured as follows. Next section introduces the market used in the model as well as the charging schemes. Section 3 describes the model and section 4 the case study. In section 5, the results are presented. Section 6 discusses the approach and section 7 concludes.

2 Market and charging

In the modelling, we focus on energy markets similar to the Nordic European countries, Norway, Sweden, and Denmark. We include both the spot market and the regulation market. We assume that the spot price is know when planning the charging on this market. However, the regulation prices are unknown and uncertain. Hence, we are aiming to see if the increased details in modelling and also the ability to bid on the regulating market will create a value for the vehicle owners - either by bidding themselves (we are aware that this might not be a possibility, due to minimum bid sizes)

or by having an aggregator controlling a fleet of vehicle bidding into the regulating market (also called the intra-day market).

When discussing up-regulation, we believe that it is questionable whether the vehicle should be able to actually discharge. However, a great deal of the up-regulation could come from not charging when planned, hence, giving back the amount not charged yet to the power system (see [11]).

Furthermore, we assume that the vehicles are always plugged in when parked. This might be too optimistic and, hence, create too much flexibility. However, when owning a fleet, this often does not create a problem, because some of the vehicles will be plugged at each time period. We will consider this when interpreting the results.

For analysing the value of information, a number of charging strategies are analysed. We are comparing the following charging strategies:

- *Uncontrolled charging* In uncontrolled or 'dumb' charging, we assume that the EVs charge their batteries as soon as they are plugged in to the electricity grid, hence, as soon as they return from a trip. Furthermore, we assume that they always fill their batteries the same amount as they have discharged while driving (keeping the battery full), being ready for the next trip. Hence, no information except the driving pattern is needed for this type of charging. Furthermore, this strategy means that the vehicle owner will not act on the regulating market.

- *Delayed charging* As with uncontrolled charging. However, the charging is delayed from when the EVs are plugged in. In this situation, some kind of intelligence is needed, in order to delay the charge (e.g. a timer setting the time for the charge to begin). However, still no actions can be taken on the regulating market.

- *Deterministic charging* The EVs optimise their charging based on deterministic future electricity prices. We optimise the EV charging based on a forecast of future market prices on the regulating market. The forecast is based on historical data of, e.g. a 1 year period. Price variations are, over a longer period, assumed to be similar, thus, a charging strategy based on these could add further value to the vehicle owner. Hence, in this situation the vehicle owners can place bids on the regulating market, increasing the value of the EV. It is assumed that the vehicle owners can act both on the up- and down-regulating market. However, up-regulation can only be done in terms of stopping or downscaling already planned demand (that is, no discharging of the vehicles are performed).

- *Stochastic charging* As with the deterministic charging, the EVs optimise their charging based on expected future electricity prices. However, here the market prices on the regulating market are considered uncertain, and, thus, they will be based on probabilities of future prices going up or down. This is done, using a two stage stochastic optimisation model. A number of scenarios will be developed, to represent different possible price development paths. Charging decisions are made before realising the actual regulating price. Here, we are increasing the details of information in terms of the variation in historical price developments. This will be based on stochastic optimisation and as with the deterministic model, this enables bidding on the regulating market.

Our hypothesis is that increasing the details in the modelling the charging decisions will also increase the benefits for the vehicle owner (and decrease the costs of electricity). However, the question is to which extend and, hence, how advanced the decision support system needs to be for the vehicle owner to benefit from these. Furthermore, the extent to which the vehicle owner can

play an active role in the power system with benefits is also expected to increase with increased information.

We are aware, that in order to place bids on the market, the bids need to have a certain size. This analysis shows the benefits for single vehicles. Combining these in different ways in order to meet the magnitude will then give different benefits. Hence, the analyses can also give an aggregator an indication of which vehicle types or combinations of these are more attractive than others.

3 Modelling description

In the following, we are assuming the vehicle owner is a price taker. We are focusing on one operation day - a 24-hour time period. This will be divided into 24 time steps, where t=1 represents the first hour, thus, the one between 00:00 and 01:00. t=0 represents the time period before the calculation period.

3.1 Uncontrolled charging

The charging after each trip can be calculated using the following formula.

$$Ch_t = min\{Ch^{max}; \sum_{k=t-\tau-\mu}^{t-\mu} Dr_k - \eta \cdot \sum_{k=t-\mu}^{t-1} Ch_k\} \quad (1)$$

Where, Ch_t is the planned charging at time t in the spot market, Ch^{max} is the maximum charging within each time step, and Dr_t is the driving at time t. τ is the length of the trip, μ is the number of time steps the vehicle has been charging continuously, and η is the charging efficiency. The equation reflects the fact that if the vehicle has used more power on the trip than can be charged within the first hour (due to grid connection), the charging continues in the next hours, until fully charged.

Based on the above, the costs can be calculated by;

$$z = \sum_{t=1}^{T} P_t^{spot} \cdot Ch_t \quad (2)$$

Where P_t^{spot} is the spot price.

3.2 Delayed charging

As with uncontrolled charging, this can be calculated by multiplying spot price and charging. We are assuming that the vehicles are charging at night, whenever possible. Now, the charging equation will be;

$$Ch_t = min\{Ch^{max}; \sum_{k=t-24-\mu}^{t-\mu} Dr_k - \eta \cdot \sum_{k=t-\mu}^{t-1} Ch_k\} \quad (3)$$

Hence, the driving from the past 24 hours is summed, and the vehicle is charged to be able to meet the next 24 hours. This equation hold from the starting time, e.g. midnight, and until the vehicle is fully charged. Then the charging starts over 24 hours later, e.g. at midnight.

3.3 Deterministic charging

For deterministic charging, we are minimising the costs of charging the vehicles.

$$\min z \ = \sum_{t=1}^{T} (P_t^{spot} \cdot Ch_t + (PE_t^{down} \cdot \lambda_t^{down} - PE_t^{up} \cdot \lambda_t^{up})) \tag{4}$$

Where PE_t^{up} and PE_t^{down} are the expected up and down regulation prices respectively . λ_t^{up} and λ_t^{down} is the charging in the regulating market.

Storage, St_t, is balanced in each time period in order to meet restrictions on storage capacity as well as the need for driving:

$$\begin{aligned} &St_t = St_{t-1} + \eta \cdot Ch_t + \eta \cdot \lambda_t^{down} - \eta \cdot \lambda_t^{up} - Dr_t, \ \ \forall t = 1, ..., T \\ &St^{min} \leq St_t \leq St^{max}, \ \ t = 1, ..., T \end{aligned} \tag{5}$$

Charging has to be within the grid capacities;

$$Ch_t \leq Ch^{max} \tag{6}$$

Furthermore, restrictions are made in order to ensure, that driving and charging cannot happen at the same time.

$$\begin{aligned} &Ch_t \cdot Dr_t = 0, \ \ t = 1, ..., T \\ &\lambda_t^{up} \cdot Dr_t = 0, \ \ t = 1, ..., T \\ &\lambda_t^{down} \cdot Dr_t = 0, \ \ t = 1, ..., T \end{aligned} \tag{7}$$

Because of the assumed up-regulation not being an actual discharge of the battery, we also need to ensure that the charging is always greater than λ_t^{up}.

$$Ch_t - \lambda_t^{up} \geq 0, \ \ t = 1, ..., T \tag{8}$$

And finally, we have the non-negativity constraints:

$$St_t, Ch_t, \lambda_t^{up}, \lambda_t^{down} \geq 0, \ \ t = 1, ..., T \tag{9}$$

3.4 Stochastic charging

In stochastic charging, the regulating prices are uncertain. Compared to the deterministic model, we have introduced the scenarios, s, and probabilities for each scenario to be realised, π_s, in the stochastic model. The deterministic equivalent to the stochastic program is:

$$
\begin{aligned}
\min z \;\; &= \sum_{t=1}^{T}(P_t^{spot}\cdot(Ch_t) + \sum_{s=1}^{S}\pi_s((PE_{t,s}^{down}\cdot\lambda_{t,s}^{down} - PE_{t,s}^{up}\cdot\lambda_{t,s}^{up}))) \\
\text{s.t. } & St_{t,s} = St_{t-1,s} + \eta\cdot Ch_t + \eta\cdot\lambda_{t,s}^{down} - \eta\cdot\lambda_{t,s}^{up} - Dr_t, \;\; \forall t = 1,...,T, s = 1,...,S \\
& St^{min} \leq St_{t,s} \leq St^{max}, \;\; t = 1,...,T, s = 1,...,S \\
& Ch_t \leq Ch^{max}, \;\; t = 1,...,T \\
& Ch_t \cdot Dr_t = 0, \;\; t = 1,...,T \\
& \lambda_{t,s}^{up}\cdot Dr_t = 0, \;\; t = 1,...,T, s = 1,...,S \\
& \lambda_{t,s}^{down}\cdot Dr_t = 0, \;\; t = 1,...,T, s = 1,...,S \\
& Ch_t - \lambda_{t,s}^{up} \geq 0, \;\; t = 1,...,T, s = 1,...,S \\
& St_{t,s}, \lambda_{t,s}^{up}, \lambda_{t,s}^{down} \geq 0, \;\; t = 1,...,T, s = 1,...,S \\
& Ch_t \geq 0, \;\; t = 1,...,T
\end{aligned}
\tag{10}
$$

Furthermore, we have introduced a constraint saying that you cannot provide up regulation if down regulation is needed and vice versa. This was needed, since in some cases it could pay of to plan charging and then provide up-regulation in these hours later - even though some scenarios were generating worse prices.

As can be seen from the model, the second stage decision (the up and down regulation) is decided upon based on a span of future regulating prices, and the first stage decision, hence the charging in the spot market, is based on the specific realization of up and down regulation.

Scenario generation will be described in section 4.4.

4 Case study

Analysing the value of added information, we focus on one vehicle type; a Nissan Leaf. Specifications are given in Table 1. Nissan Leaf has two different battery use settings; long distance using the battery 100% or long life using the battery 80%[9]. We are using the long distance and, hence, assuming that 100% of the battery is available for driving and charging. However, in our analyses, the battery is never depleted below 20%. Thus, we might as well use the long life.

Parameter	unit	value
Battery capacity	kWh	24
Efficiency	km/kWh	5.8
Total charging time	hours	6 - 7
Max driving per charge	km	199

Table 1: Based on ([5])

4.1 Data and assumptions

We are assuming the vehicles are plugged to the electricity grid whenever they are parked. Each vehicle have an assumed connection with 3 phases 10 Amps, resulting in a grid connection capacity

of 6.9 kW. Hence, maximum charging capacity in each hour is 6.9 kWh. Furthermore, we have assumed a charging efficiency of 0.9.

4.2 Driving patterns

We use the clustered driving patterns found in [14]. All 20 patterns are included in order to get an idea whether there are driving patterns or life styles where it is of greater value with more sophisticated modelling than others.

This way we can also analyse whether it is more beneficial to own a fleet of vehicles with different driving patterns or a fleet with the same driving patterns.

4.3 Spot prices

For our analyses, we have used historical hourly spot prices from four days in four different seasons in 2014, hence, January 1^{st}, April 1^{st}, July 1^{st}, and October 1^{st}.

4.4 Regulation prices

Forecasting differs for the different analyses. No forecasting is needed for uncontrolled and delayed charging. For the deterministic model, we are using an average of the price deviation from spot to regulation prices on an hourly basis, based on data from year 2013. This corresponds to the average of the scenarios for the stochastic analysis. Hence, the regulating price is calculated based on the spot price plus/minus the deviation (depending on whether it is up or down regulation).

4.4.1 Scenario generation

Scenarios are based on data from year 2013. Regulation price scenarios are generated by means of the heuristic method described in [12]. The regulation price at each hour of the day is modelled as an independent random variable. The method uses marginal distributions for the random variables and copulas to describe the dependence between the marginal distributions. Marginal distributions and copulas have been estimated based on historical regulation prices.

5 Results

Results show that we experience decreasing costs with increasing intelligence in the charging decision. Figure 1 shows the total costs of charging the 20 different vehicle types (one of each). As seen from the figure, a large decrease is experienced between uncontrolled and delayed charging, hence, only moving charging to the night time. However, another large decrease can be found using either deterministic or stochastic modelling, especially with the electricity prices in the April data.

As for the gain of using stochastic modelling instead of deterministic, we use the value of the stochastic solution (VSS) as found in [3]:

$$VSS = E[z(x(E\xi), \xi)] - E[z(x^*, \xi)] \tag{11}$$

However, when we try to solve the stochastic problem using the first stage solution of the deterministic, the problem becomes infeasible. This has been tried both with the implemented scenarios and another set of scenarios. Infeasibility of $E[z(x^*, \xi)]$ is equivalent to a very high VSS. Hence, the solution to the deterministic problem is not robust towards slight changes in the regulation prices and, thus, regulation possibilities.

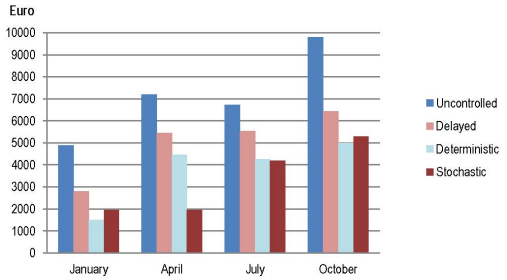

Figure 1: Total costs for different charging schemes

This lack of robustness is partially due to some inappropriate planning on the up-regulation side. If we try to remove the possibility of up-regulation, we get the following VSS for the total of all 20 vehicle types:

January	9.762 €
April	613.594 €
July	584.433 €
October	145.585 €

The rather high values for April and July, is because the stochastic solution only charges the vehicles on the regulating market. Hence, we count on the need for enough down regulation at some point during the day, when the car is parked. Looking at the charging pattern as well as up and down regulation, it is evident that almost all of it is in the night time. Hence, most of the vehicles will be parked and the assumption that vehicles are plugged in when not driving, does not influence our results much if at all.

In Figure 2, we see the cost average from the four different seasons using the different clusters of driving patterns. From the figure we can see that we experience a decrease in costs between 40-60% when using the stochastic solution. Furthermore, focusing on Figure 3 we see that the monetary saving is quite different for the different vehicle types - both because of the different charging needs, but also due to the different timing opportunities for charging.

Based on these analyses, we see that the savings for vehicle types EV10, EV16, EV19, and EV20 are very low. If an aggregator only has these vehicle groups in his fleet, the increment to act smart on the regulating market is a lot smaller than with a fleet of, e.g. EV11.

6 Discussion

From the results we see that in general it is of great value to introduce a stochastic model to optimise the charging and bidding on the regulation market for electric vehicles. The results could be scaled to a large number of vehicles, imitating that of an aggregator. However, we need to keep in mind that an increased number of vehicles does not increase the expected savings proportionally. The regulating market only needs a certain amount of regulating power. However, having a diversified fleet could enable you to bid into the market at most hours and almost always have available

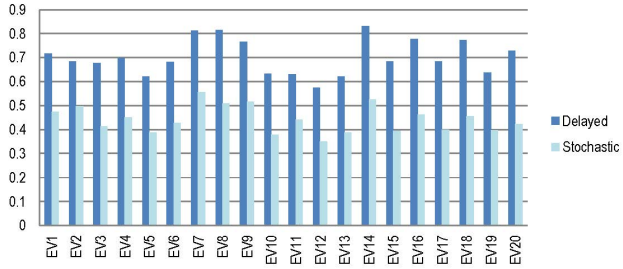

Figure 2: Average costs as percentage of cost of uncontrolled charging

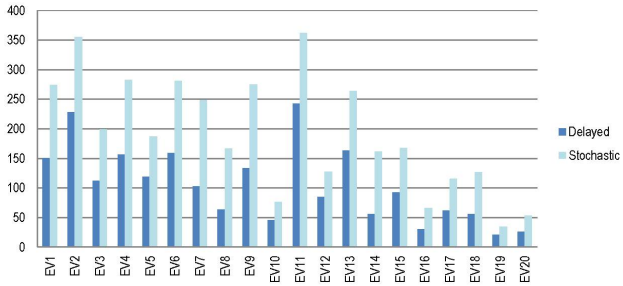

Figure 3: Average costs savings (€)

vehicles, increasing the expected earnings.

The model developed in this article could be enhanced to either include rolling planning for more details (up to the 24 hourly time steps included, if to compare), or to develop a multi-stage stochastic model. Hereby, the value of more sophisticated modelling could be studied as well as to which extend it is still beneficial to increase the details. Could it, e.g. be optimal to split the 24 hour day into 4 stages and model this, since making a 24-stage model easily make the number of scenarios explode?

Furthermore, one could argue that normally reserves are not needed in the same direction throughout the complete hour. However, the intra-day market works on an hourly basis, but often does not allow for us to, e.g. provide down regulation services for the complete hour. Adjusting

the modelling to take the uncertainty of the amount of power to be available for regulation could, therefore, also be a subject of further research.

Finally, other devices in the power system can also provide the same kind of demand response as the EVs and could easily benefit from the detailed modelling as well. It could, e.g. be interesting to look into the values for electric heating, electric boilers, and electric cooling.

7 Conclusion

Using mathematical models for charging the electric vehicles adds value to the vehicle owners or aggregators. The value varies between the different uses of the vehicles, for some the value is large, for others, the planning most likely does not give a value great enough for one actor to consider pooling with others and implementing the necessary intelligence in the vehicle.

We have showed, that for acting on the regulating market, the value of a stochastic model over a deterministic model is very high (with an infeasible stochastic solution to the deterministic first stage). Only focusing on the Day-Ahead market with possibilities for down regulation, also results in a rather large VSS. Moving on to more detailed stochastic models might increase the value even further.

8 Acknowledgements

This article is part of the ENSYMORA project funded by the Danish strategic research council.

9 References

References

[1] Al-Awami AT, Sortomme E, Coordinating Vehicle-to-Grid Services With Energy Trading, IEEE Transactions on smart grid 3(1)(2012), pages 453-462. http://dx.doi.org/10.1109/TSG.2011.2167992

[2] Bessa RJ, Matos MA, Optimization models for an EV aggregator selling secondary reserve in the electricity market, Electric Power Systems Research 106 (2014), pages 36-50. http://dx.doi.org/10.1016/j.epsr.2013.08.006

[3] Birge JR, Louveaux F, Introduction to stochastic programming, New York: Springer; 1997

[4] Boomsma TK, Juul N, and Fleten SE, Bidding in sequential electricity markets: The Nordic case, European Journal of Operations Research 238 (2014), pages 797-809. http://dx.doi.org/10.1016/j.ejor.2014.04.027

[5] Clever, https://www.clever.dk/produkter/faa-1-krone-tilbage-pr-kwh-stroem-du-oplader-elbilen-med/

[6] Dallinger D, Krampe D, Wietschel M, Vehicle-to-grid Regulation Reserves Based on a Dynamic Simulation of Mobility Behavior, IEEE Transactions on Smart Grids 2(2)(2011), pages 1949-3053. http://dx.doi.org/10.1109/TSG.2011.2131692

[7] Fleten SE, Kristoffersen TK, Stochastic programming for optimizing bidding strategies of a Nordic hydropower producer, European Journal of Operational Research 18(2)(2007), pages 916928. http://dx.doi.org/10.1016/j.ejor.2006.08.023

[8] Foster JM, Caramanis MC, Energy Reserves and Clearing in Stochastic Power Markets: The Case of Plug-in-Hybrid Vehicle Battery Charging, IEEE Conference on Decision and Control (2010), pages 1037-1044. http://dx.doi.org/10.1109/cdc.2010.5717304

[9] Green autoblog, http://green.autoblog.com/2013/02/21/2013-nissan-leaf-revealed-gets-75-mile-range-actually-84-in-n/

[10] Iversen JEB, Madsen H, Morales Gonzlez JM, Optimal charging of an electric vehicle using a Markov decision process, Applied Energy 123 (2014), pages 1-12. http://dx.doi.org/10.1016/j.apenergy.2014.02.003

[11] Juul N, Meibom P, Optimal configuration of an integrated power and transport system, Energy 36(5)(2011), pages 3523-3530. http://dx.doi.org/10.1016/j.energy.2011.03.058

[12] Kaut M, A copula-based heuristic for scenario generation, Computational Management Science 11(4)(2014), pages 503-516. http://dx.doi.org/10.1007/s10287-013-0184-4

[13] Khalid MW, Al-Awami AT, Sortomme E, Stochastic-Programming-Based Bidding Strategy for V2G Services, 4th IEEE PES Smart Grid Technologies Europe (2013), pages 1-5 http://dx.doi.org/10.1109/isgteurope.2013.6695468

[14] Kristoffersen TK, Capion KE, and Meibom P, Optimal charging of electric drive vehicles in a market environment, Applied Energy 88(5)(2011), pages 1940-1948. http://dx.doi.org/10.1016/j.apenergy.2010.12.015

[15] Ladurantaye D, de Gendreau M, and Potvin JY, Strategic bidding for price-taker hydroelectricity producers, IEEE Transactions on Power Systems 22(4)(2007), pages 21872203. http://dx.doi.org/10.1109/TPWRS.2007.907457

[16] Lund H, Kempton W, Integration of renewable energy into the transport and electricity sectors through V2G, Energy Policy 36(2008), pages 35783587. http://dx.doi.org/10.1016/j.enpol.2008.06.007

[17] Vagropoulos SI, Bakirtzis AG, Optimal Bidding Strategy for Electric Vehicle Aggregators in Electricity Markets, IEEE Transactions on Power Systems 28(4)(2013), pages 4031-4041. http://dx.doi.org/10.1109/TPWRS.2013.2274673

[This page intentionally left blank]

[This page intentionally left blank]